

Mikhail Gorbachev is back

The pre-winter equinox is a time for getting back to serious business. Mr Gorbachev and his counterparts in the West have a great deal to do before they let the northern winter take over.

MR Mikhail Gorbachev's return from his annual summer vacation is fast becoming an annual global catharsis. When he is away, Moscow is filled with dark rumours and the mutterings now allowed by *glasnost*, while the rest of the world acts as if it has been unsettled. Then he comes back. Sometimes (as two years ago), he announces that he has written a book. On other occasions, as last week, he contrives confidently to sound as if insoluble problems are also the only problems with which it is worth bothering. Russia would not be the driving force in the Soviet Union if there were not an infinity of insoluble problems, but some, even so, are conspicuous.

There is, for example, what is called the ethnic problem, about which Gorbachev would appear to his predecessors (and does to his contemporary conservative opponents) to be curiously indifferent. Hungary plans to follow Poland, in the autumn, with a kind of free election, the three Baltic states behave as if they wish they could be up in arms, there is discontent throughout the southern republics and Moldavia would reunite with Romania if only that were congenial, but meanwhile demands the right to speak Romanian. The difficulty, for the Soviet republics, is that there are too many of them to cut much ice. Their friends elsewhere are mostly remittance men. In due course, it may be different, but many other things will by then have changed the world.

Second, there is the problem of the Soviet economy, which is bankrupt. The Soviet people cannot feed, or clothe or adequately house themselves. Gorbachev has railed against Stalin as energetically as anybody in his position could, but he has not yet spelled out what needs most simply to be said: that Stalin's exercise of his unchallenged right to squander half a century's investment of the Soviet people's enthusiasm on misguided enterprises is a more serious offence than his better-documented misdemeanours. Yet that is where the truth lies. Nobody asks that the Soviet people should repudiate the past half century, or even pretend that it never happened, but merely that Gorbachev, back from his vacation, should refer to the difficulty from time to time, by way of explanation of why that old enthusiasm has mostly been nugatory.

Third, there is the ideological problem. Gorbachev's licence as president of the Soviet Union is that of a leninist — the right to run a government on marxist lines. If the rules were different, he might have to apply for another

licence; nobody can be sure what the outcome of that would be. The unspoken question at Gorbachev's return from his summer vacation is that the unspoken disloyal ideological question may now be unavoidable. If that were indeed the case, of course, everything would change. Gorbachev would probably retire, or take a post as president of a prosperous university in the West. It would be convenient for us all if that issue could be postponed for another year.

So what happens, or should happen, as the long winter closes in on the Soviet Union? First, there needs to be a recognition in the West that, of all the crises in the world, that in the Soviet Union should command the most attention. It is, after all, the most important potential crisis. If, this winter, there should be famine in the Ural district or somewhere else, there will be another Western-looking Dostoevsky, or perhaps several, on the West's back for the rest of time. Second, the outside world (not necessarily coextensive with the West) needs some kind of emergency plan. Third, quite soon, the US administration needs to settle on a plan for arms-control; Washington professes to be preoccupied with schemes for winning advantages from the present hiatus: why does it not offer a test-ban treaty (virtually completed nine years ago) as a gambit, knowing that there can hardly be a more secure agreement? □

No news is bad news

Impatience and intemperance may have persuaded the British government to cancel a timely project.

SQUEAMISHNESS is not one of the qualities on which the British government of the past decade prides itself. On the contrary, the government boasts (and even boasts about) qualities at the other end of the spectrum, tough-mindedness for example. Yet Mrs Margaret Thatcher, the prime minister, has now used her considerable influence to stifle at birth a research programme meant to help understand the mechanism of the spread of AIDS in Britain. The objective is a social survey of contemporary adult behaviour, sexual behaviour included. The prime minister's office says that she considers that the study would have been "too intrusive" to be supported with public funds. Thatcher may be embarrassed to think of



20,000 of her electors answering questions she might prefer not to be asked herself, or she may fear for the consequences on the same respondents' voting behaviour. Either way, it is squeamish.

And worse. The survey now stopped in its tracks has been in gestation for the past two years. A large group of British AIDS researchers has devised not merely a list of questions to be asked, but methods by which they could be asked unobtrusively. A random sample of people (chosen from the Post Office Small Users' List) would have been invited to participate, but would have been free to refuse. The questionnaire would have been 'self-completed' in the argot of the social science trade: respondents would have completed the questionnaire themselves, sealing it into an envelope provided, referring to the survey official delivering the questionnaire only in case of ambiguity. Two pilot surveys, with 2,000 and 5,000 people respectively, have shown that the response is good and complaints a mere trickle. Arbitrarily calling a halt to such a careful project is a slap in the face for the whole of British science.

But it is worse than that. The titularly autonomous Economic and Social Research Council (ESRC), the most obvious British backer, approved the final version of the proposal as long ago as February, promising a contractual commitment within the week. Because its own budget is only meagre, ESRC had by then recruited the Health Education Authority and the Department of Health as co-subscribers to the estimated cost of £510,000; plainly the council had an obligation to consult the bigger money-bags, but equally plainly it had cause to believe that agreement would be a formality, both partners having been fully consulted along the way. But that week and about 30 more came and went with nothing happening — except that the Health Education Authority at one stage decided not to continue paying for the cost of the second pilot study it had agreed to share. Letters to ministers (one from a House of Commons select committee) went unacknowledged. Even a parliamentary question was somehow overlooked. Maladministration is the polite name for muddle of that kind, but there is squeamishness as well. Even if the prime minister's colleagues had weakly passed the buck to her, it was squeamish of them to let last week's decision be so long delayed.

Worse still, this arbitrary decision will damage the chances that AIDS will be rationally managed in Britain. First recognized eight years ago, AIDS is still an emerging scourge. That so much has been learned so quickly about the mechanism of infection by the human immunodeficiency virus (HIV) and its palliation may be remarkable, but it remains an open question whether AIDS will be contained within the groups it now predominantly afflicts (homosexual and bisexual men, intravenous drug users and the recipients of blood transfusions) or will spread more widely, especially among heterosexuals. In the widely accepted terminology of Anderson and May (*Nature* 333, 514; 1988), the crucial parameters are the

likelihood that an infected person will infect a sexual partner and the rate at which people acquire new sexual partners. As drugs such as AZT are more widely used as palliatives of overt and (now) cryptic AIDS, the infectivity of people carrying HIV at different stages will also become important. Social surveys of the kind now planned are the only routes to understanding. By intemperately saying "No", Thatcher has denied her fellow-citizens both a basis for telling how widely AIDS will spread and — ironically — an incentive to mend their behaviour in prudent ways. □

A new beginning?

The British Association's stirring manifesto from Sheffield would be more compelling if it were more immediate.

THE British Association for the Advancement of Science, widely regarded as moribund, has managed to organize a second successful annual meeting in a row. This year, at Sheffield, the association has even issued a "Charter for Science", pleading for a more deliberate approach to the revivification of British science. The association and the distinguished signatories of its charter are of course right to say that the new effort should begin in the schools: where else? They are neglectful in not asking why it should be necessary to begin again, when only a few years ago it seemed as if other beginnings had matured into success. Good manners notwithstanding, it now makes very little sense that the distinguished survivors from a distinguished tradition should urge their fellow-taxpayers that if only more attention were paid to the quality and character of secondary education, Britain might yet recreate a Newton or a Darwin (neither of whom won a Nobel prize — they were injudiciously precocious). A little anger would have been more appropriate.

The truth about British science is that it has been made to serve as the living proof that it takes only a few minutes to get rid of something whose survival depends on the quality of people's spirits: the mechanism is simply to make them depressed. In Britain, the process has been sustained for the best part of a quarter of a century, but the demands of the past decade have been unnaturally cruel. University departments have been robbed of staff, research councils (while their budgets have been protected against inflation) have been told how to spend their money on wealth-creating projects and often, in their anxiety to please, have tried unsuccessfully to second-guess their sponsors (witness present official discontent with the practice, as distinct from the principle, of interdisciplinary research centres). The charter for science would have been more compelling if fewer of its signatories had been (or were still) party to the calamitous decisions of the past few years.

What is to be done? The British Association is right to say that it would be best to begin again in the schools, but even if that were feasible, could Britain (or any other

second-rank power) afford to wait that long? Another couple of decades, perhaps? Of course not. The effort in the schools to which the British Association would dedicate itself is essential, but is a long-term project. The immediate need is to find a way of helping researchers now in post to believe cheerfully in their work, even when the prospect that they will earn substantial royalty fees from industry is small. Maybe, the British prime minister's decision on the AIDS survey notwithstanding, the British Association will find an answer by next year. □

More funny money

The only solution of Europe's money problem is to require European states to accept ECUs as legal tender.

MR Nigel Lawson, the British chancellor of the exchequer, is in a curious dilemma. Last week, which he spent at Antibes in the company of other finance ministers of the 12 members of the European Community, he argued vigorously (as is his wont) against the notion that there should be a common European currency and a central bank to go with it. But next weekend he will be in the United States, arguing that Britain should remain the second largest subscriber to the International Monetary Fund, which is a perfectly respectable international central bank. Why the contradiction?

The puzzle is most easily explained by supposing that what Lawson said last weekend at Antibes was a kind of joke. Faced with the job (for which he volunteered at Madrid in June) of devising alternatives to Europe's plan for a common currency, he appears to have offered the notion that all European currencies should be considered equal, coupled with the hope that some in due course would come to seem more equal than others. Lawson appears not to have committed his plan to paper, so that it is not certain what he meant. That may not matter much, given that critics of the plan (among whom Lawson may be included) are unlikely to give it an easy passage.

But there is a need for a European currency. Otherwise the banks will be the only winners from the single market. But at this late stage in history, it can make no sense to devise some non-monetary substitute for a currency — germanium, or even palladium, would not suit. Notions that the commodity called energy would make the basis for the ideal unit of currency are of course mistaken; if currencies must be linked with commodities, the commodities must be otherwise useless, as gold used to be.

Nor can it make sense that there should be a kind of competition between national currencies, as Lawson proposes. The optimists may believe that good currencies will take precedence, but will it not still be advantageous to pay for goods and services in the cheapest, not the most expensive currency, thus verifying Gresham's Law The obvious solution is to have two legal currencies everywhere — the national currency and the European Currency Unit. Then you feel your way. □

End of apartheid?

The South African election last week must be demonstrated to have been a hopeful sign or it will be the last.

WHITE South Africa, which re-elected its Nationalist government last week with a reduced majority, has a nasty job to tackle next: it has to persuade its new government, elected on the basis of "moderate reform", that there is no such thing. Reform is either radical — of or pertaining to the root — or it is not reform. South Africa's immediate problem is that its new government, having survived an election marked by brutality against non-whites, unjustifiable even by the calculation that its performance would otherwise have been even worse, and the opposition more wedded to *apartheid*, may still be preoccupied with electoral arithmetic. The government must be made to understand that it is unique among elected governments in being the one whose success will hang on whether it can create the conditions in which there will be another election of any kind by which its successor can be legitimized.

The article on page 96 by Dr Stuart John Saunders, the vice-chancellor of the University of Cape Town, and his colleagues typifies the process of persuasion that must be begun. Neither readers of this journal nor the government of South Africa should be lulled into complacency by the simplicity of what they have to say. It will be noted, for example, that Saunders *et al.* use the word "post-apartheid" often, especially towards the end of their article. What does that mean? Simply, it means that they believe there will be a time when *apartheid* has disappeared, and when the need to provide secondary schooling for those whom white South Africans now constitutionally set apart will be so great that even traditionally proud universities may have to reset their sights at more practical goals than those of excellence. The comfort, for Saunders and his colleagues and the rest of us, is that they are by no means alone.

Mr F.W. de Klerk, this week's new state president of South Africa, is strangely more isolated. His difficulty may no longer be electoral, but he has assiduously sought (and won) the sufferance of people outside South Africa (in Zambia as well as Western Europe) by his assurances that reform means the root-and-branch stuff. If his election last week had passed off peacefully, he might have had some time in which to demonstrate that he meant what he said. As things are, there is less time — indeed, there is very little. The release of Nelson Mandela is at the top of every reformer's list, but should be a formality. Abolishing the Group Areas Act, or something like it, by the end of the month would be more to the point. But the enduring need is to find the funds with which to educate all South Africans. If de Klerk does not hurry in that direction, others elsewhere should take on the burden. Whatever happens, people will be keeping a close eye on South Africa. □

New York State leads on genetic fingerprinting

- First attempt at regulation
- New labs and databank proposed

Albany, New York

The first attempt in the United States to regulate the forensic use of 'genetic fingerprinting' was announced last week by the director of justice for New York state, John Poklemba. According to a plan likely to be approved and implemented next year, the state would establish a committee to set and monitor standards, and a scientific board to examine the accuracy of test results before they could be presented in court.

There is at present no federal or state legislation regulating DNA testing in private or public laboratories. Although the need for some kind of monitoring is widely acknowledged, there is much dispute over who should do it. And the dispute will grow hotter as more state forensic laboratories carry out DNA testing.

Only one of the nearly 300 forensic laboratories in the United States now does DNA testing, according to a survey carried out by the Congressional Office of Technology Assessment (OTA), but a draft OTA report on the forensic use of DNA tests says that interest has "skyrocketed" in the past two years, and that most laboratories have plans to start testing in the next two years.

The report also warns that if no national regulatory action is taken, "continued case-by-case examination of proper tech-

SPACE

Japanese success

Tokyo

OFFICIALS of Japan's National Space Development Agency (NASDA) last week breathed a sigh of relief as an H-1 rocket successfully carried a meteorological satellite (GMS-4) into orbit. An attempt to launch the satellite a month ago failed because the first stage of the H-1 rocket did not fully ignite (*Nature* 340, 493; 17 August 1989).

There were fears that NASDA might have to postpone the launch until next year because launches from NASDA's space centre on Tanegashima island base are limited to particular periods by an agreement with local fishermen. A large sea area close to the base has to be cleared before each launch. The satellite, named Himawari (Sunflower) No. 4 will be moved to a geostationary orbit in about a month and is expected to provide meteorological data for the western Pacific region for five years starting in mid-December.

David Swinbanks

nical and operational standards could slow full implementation of forensic analysis using DNA typing".

The Federal Bureau of Investigation (FBI), a leader in forensic DNA analysis, is holding a series of meetings with researchers and laboratory representatives aimed at reaching a consensus on technical standards. The group is said to be close to agreement on the main issues, and next month's meeting may be the last.

But there are objections to the leading role of the FBI, an investigative and law enforcement body, in setting standards.

Next month, the National Academy of Sciences will begin a comprehensive study of the technical, legal and ethical issues involved in the forensic use of DNA analysis, having obtained promises of funds from a variety of sources including the FBI, the National Institute of Justice and the National Institutes of Health. A report is expected in late 1990.

The regulations proposed for New York have been warmly welcomed by researchers and lawyers grappling with the problems created by the transfer of the techniques of DNA analysis to the courtroom. Peter Neufeld, a lawyer on the panel that produced the plan, said they would have prevented some of the problems that surfaced in the recent New York double-murder trial (see *Nature* 340, 582; 24 August 1989). And Eric Lander, of the Whitehead Institute for Biomedical Research, who was an expert witness at the trial, called the plan "fabulous".

Among its recommendations is a proposal for at least three new state-run laboratories. According to the New York panel, "questions about the quality of the work being done by the private laboratories have not been satisfactorily answered, and the laboratories' adherence to accepted scientific procedures has not been demonstrated". It says that "sweeping claims of accuracy, stating that the probability of error is one in a million, or in some cases one in a billion" are "suspect". And it criticizes the laboratories for being "reluctant" to share information about their procedures. If DNA testing results are to be accepted as evidence in courts, the techniques must be generally accepted as reliable in the scientific community. But the panel says that "it is difficult to reconcile the practice of cloaking a methodology in secrecy with the claim that the methodology is widely accepted".

The DNA analyses carried out by the state laboratories, expected to cost about \$500,000 to set up, would be equally available to defence and prosecution and the costs would be apportioned by some mechanism other than on a per-case basis. "Decisions on whether DNA analysis will be applied in a given case should be made on the merits of the case, not on whether there is sufficient money in the budget to pay for the analysis", says the panel. And in order to generate population statistics, all data generated should be kept but should not contain information traceable to an individual.

According to OTA, resources for training forensic laboratory staff are "woefully inadequate". At a meeting of the OTA advisory panel last week, chairman Thomas Caskey of Baylor College of Medicine said that the United States is lagging behind the United Kingdom in the forensic application of DNA analyses and more funds for training are needed. He also said a committee should be formed to set national guidelines on quality assurance.

Under the New York plan, private laboratories could still perform DNA testing, but would have to be accredited by the state; this would require independent validation of the methods used and participation in state proficiency testing programmes. Approval by a scientific review board, which would set minimum scientific controls and examine the tests used, would be necessary before the results could be presented in court. The panel says that the usual method of determining whether a technique is generally accepted in the scientific community, by publication in peer-reviewed journals, is "inappropriate and unrealistic" in the case of DNA analyses in forensic work.

The panel also recommends the creation of a DNA databank, provided there are regulations to prevent use of the databank for any purpose other than identification of suspected criminals.

Only a computerized code allowing identification would be stored, and a match should not be the sole basis for making an arrest.

This strikes "an appropriate balance between competing privacy and legitimate law enforcement interests", says the panel. It also strongly urges the establishment of national standards for all testing procedures, analysis, interpretation and coding of data so that databases can be shared as well as quality controls to be carried out nationally. The FBI is due later this year to release a report on the establishment of a national database.

Several states have already enacted legislation that requires DNA testing of convicted sex offenders, and legislation is under consideration in others. But no state has yet established a databank.

Christine McGourty

Space science is a winner

Tokyo

JAPAN's space scientists and astronomers find favour in the 1990 budget request just submitted by the Ministry of Education, Science and Culture (MESC) to the Ministry of Finance. Although total funds for space science are reduced, the budget is full of small but significant requests for 'seed money' to begin major new projects in space science and astronomy. Global environment research also gets a boost.

The ministry's budget will not be finalized until the end of the year but is likely to pass the Ministry of Finance without major amendments.

One of the most significant items in the budget is a tiny request of ¥14 million (\$100,000) in preparatory funds for a 7.5-metre infrared telescope to be built at Mauna Kea in Hawaii by the National Astronomical Observatory. The observatory, which until recently was part of

1980 Science budget request for Ministry of Education, Science and Culture

	thousand million yen	(% change from 1989)
Grants-in-aid of research	56.0	(+ 6.4)
Government/Industry research	11.4	(+10.7)
Donations from Industry	39.2	(+18.0)
Research fellowships	2.2	(+14.1)
Nuclear fusion	9.0	(+ 4.8)
Accelerator physics (TRISTAN)	16.4	(+ 3.2)
Space science	18.3	(-12.3)
Astronomy	0.9	—
Environment	4.0	(+32.2)
Earth Science	2.3	(+ 8.8)
Antarctic research	5.1	(+73.1)
International exchange	6.6	(+12.4)

Tokyo University, has been trying to win support for the telescope for years (see *Nature*, 311, 5; 1984).

The telescope will have a very high resolving power (about 0.1 arc seconds) and wide field of view, making it ideal for the study of the processes of formation of stars, planets and galaxies. The total cost of the telescope and related facilities is expected to be about ¥45,000 million (\$312 million), more than twice the amount estimated in 1984 when plans for the telescope were first made public.

Also hidden away in the budget request is a small allocation for the development of a new solid fuel rocket for the Institute of Space and Astronautical Science (ISAS). The rocket will have about three times the power of the institute's present biggest rocket the MU-3SII and will cost about \$150 million to develop. One of its first missions will be to launch a radio-telescope satellite (see page 92). And the budget for space science will rise after

1990 as development costs of the rocket and satellite increase.

Radioastronomy gets another big boost with a request for ¥845 million (\$6 million) to build a heliograph at the Nobeyama Radio Observatory of the National Astronomical Observatory. The heliograph will consist of a T-shaped array of 76 80-cm parabolic antennas. It will be completed in two years at a total cost of about ¥1,900 million (\$13 million) and will observe solar flares during the next solar maximum in conjunction with the solar observation satellite Solar-A, which is scheduled to be launched by ISAS in 1991.

The global environment features prominently in MESC's budget request as it did in the already-released budget requests of other government ministries and agencies (see *Nature* 341, 4; 1989). And, similarly, some of the money is to be used to establish new research centres.

The Research Institute of Atmospherics at Nagoya University will be reorganized into an institute provisionally named the Solar Earth Environment Research Institute at a cost of \$1 million (¥136 million). The institute will study the inflow and outflow of energy and materials from the Earth and Sun and the effect of these fluxes on the global environment.

A similar amount (¥224 million) is requested to reorganize and rename the Sand Dune Research Institute of the University of Tottori. The revamped institute will broaden its horizons from the study of local sand dunes to research on the deserts of the world. And ¥26 million is requested to establish an Arctic research centre at the National Institute of Polar Research.

Funds for joint international research on the global environment rise 10 per cent above this year's level to ¥3,239 million (\$22.5 million). Extra backing goes to joint research on the Arctic environment in collaboration with the United States, Norway and other countries. Antarctic research also receives a huge increase (73 per cent) to ¥5,069 million (\$35 million) but much of the extra money is a one-time allocation for the purchase of a ship-borne helicopter.

David Swinbanks

■ "DONATIONS from industry", one of the biggest items in the budget request, are funds donated under a scheme established by MESC in 1983. Such donations are the fastest growing source of support for university researchers and are expected to reach nearly ¥40,000 million (\$280 million) in fiscal 1990. The government's contribution — ¥11,400 million (\$80 million) in fiscal 1990 ("government/industry research" in table) — has also grown rapidly, and the total budget for the scheme now almost matches MESC's budget for research grants. □

UK physicians demand action

London

ADVANCES in methods for genetic screening and prenatal diagnosis should be available to all couples wishing to have children, the British Royal College of Physicians (RCP) has recommended. Education about genetics should begin at school, should be a part of the medical school curriculum and should also become a part of clinical practice.

In a report published on 7 September, the RCP claims that many British couples are denied access to up-to-date screening techniques and prenatal diagnosis and that the availability of these tests varies four-fold between different health authorities, the administrative organs of the National Health Service. The report says that the test for spina bifida, available since the 1970s, is offered to only 70 per cent of women at risk, while amniocentesis for Down's syndrome is available at only 12 hospitals in Britain.

The report recommends that a nationwide screening and prenatal diagnosis scheme should be established, which would become an intrinsic part of maternal health care. Professor Martin Bobrow, a clinical geneticist at Guy's Hospital Medical School who helped to write the report, said a comprehensive screening and back-up system of diagnosis could prevent the birth of 2,000 severely handicapped babies a year. The RCP concludes, after a cost-benefit analysis of the scheme, that "if the costs of the whole programme were aggregated, it is cheaper to screen and counsel the whole population than it is to treat affected children who would otherwise be born to unprepared couples". The cost of a test is a few pounds, the cost of caring for a severely handicapped child is about £10,000 a year. Dr Margaret Turner-Warwick, RCP president, said short-term budgetary considerations meant that health authorities did not think in terms of long-term saving.

Sir Raymond Hoffenberg, a past RCP president and a member of the working party, said an integral part of the plan was to improve education. Many women were not offered tests simply because general practitioners and hospital consultants were unaware of what was available. The report says a designated co-ordinator should be responsible for organizing delivery of both specialist and community genetic services in each region.

A code of practice should be drawn up to tackle the inevitable ethical problems involved. Bobrow said prenatal diagnosis would not be used for positive eugenic policy, but to enable individuals to make informed decisions about their lives based on the fullest possible knowledge.

Ben Webb

Greens oppose EPO plant

Munich

OPponents of genetic engineering from all over West Germany descended on the sleepy university town of Marburg last week to present a wide assortment of objections to a biotechnology plant for the production of the drug erythropoietin (EPO), which stimulates the creation of red blood cells in patients with kidney failure. In protesting against the plant, the opponents, led by the Green party, took advantage of a long-awaited opportunity: the first public licensing hearing in West German history for a gene-technology facility.

The licensing battle is an important test case for the future of industrial genetic engineering in West Germany. It could affect the 'framework law' for genetic engineering scheduled to be debated in parliament this fall. Other companies, including BASF AG, are facing similar hearings in the next few months.

Both Behringwerke AG, which owns the plant, and the Greens expect the plant to be licensed despite the protest, and both expect a legal battle to follow. Behringwerke had applied to produce EPO in 1988 but submitted its application again after the law was changed to require a public hearing (see *Nature* 335, 199; 1988).

The licensing authority, which in this case is the regional government in Giessen, announced that it would rule on the application by the end of the year.

The Greens made an impressive showing at the hearing both in the volume and the content of their objections. There were nearly 2,000 written objections to the plant, many of which concerned the moral and ethical aspects of genetic engineering. But the Greens also raised technical objections that they claim Behringwerke had not been expecting.

For example, Regine Kollek of the Greens objected to the genetic vector used to transform mouse cells in culture and induce them to produce EPO. Kollek claimed that bovine papilloma virus 1 (BPV 1) is present in the cells in its entirety and that a single virus particle might cause harm to humans.

The Greens called for data on the genetic lineage of the mouse whose cells will be used for the cultures, in order to determine if the mouse carries viruses such as mouse mammary tumour virus (MMTV) that might increase the chances of recombination, with unforeseen genetic consequences.

Behringwerke responded that these charges are "scientifically irrelevant." For instance, Behringwerke denied that BPV 1 might appear in either the cells or the EPO itself. The company reported that the cell line in question has been used for

15 years without any evidence that endogenous retroviruses might be activated. But Behringwerke spokesman Wolfgang Faust said that the company expects to be asked for further documentation before a license is granted.

Marina Steindor, spokeswoman for the national Green party working group on gene technology, said the Greens made "great progress" at the hearing. Steindor said that public hearings should be a part of the licensing procedure for all biotechnology products.

In the proposed genetic engineering law, hearings would not be required for some production processes involving "harmless" strains.

The issues raised at the licensing hearing would sound familiar to those United States companies that first won permis-

RADIOASTRONOMY

Japan plans a space VLBI

Tokyo

JAPANESE scientists hope to be the first to launch into orbit a satellite that can be linked to radiotelescopes on Earth to form a huge VLBI (very-long-baseline interferometry) network.

The VLBI Space Observatory Programme (VSOP) is being promoted by the Institute of Space and Astronautical Science (ISAS) in collaboration with researchers at the Nobeyama Radio Observatory of the National Astronomical Observatory and the Communication Research Laboratory, and was described at an international symposium in Tokyo at the end of last month. A recent decision by the Space Activities Commission to allow ISAS to build a larger solid fuel rocket (*Nature* 340, 8; 6 June 1989) has given a boost to VSOP, and preliminary funds for the project are included in the budget request for fiscal year 1990 just submitted by the Ministry of Education, Science and Culture (see page 92).

ISAS hopes to launch the VLBI satellite in early 1995. Its closest competitor was the European Space Agency (ESA)'s VLBI-satellite QUASAT. But ESA cancelled the project last year, leaving only Japan and the Soviet Union in the field, given that the United States has no plans for a VLBI satellite. The Soviet Union hopes to launch the Radioastron satellite in 1993 but Japanese scientists involved in VSOP think it is likely to be delayed.

The orbit of the VSOP satellite will have an apogee of 20,000 km and, when linked to radiotelescopes on Earth, will create a VLBI radiotelescope network with an effective diameter of 30,000 km — about three times the size and resolution of the biggest networks on Earth.

sion a decade ago or more to use genetic engineering for the production of drugs. But whereas the US debate came to be dominated by pragmatism, the German debate is marked by the opposition's almost religious adherence to the belief that all genetic engineering is hazardous and should not be allowed anywhere. Those concerned seem prepared to resort to any legal means to prevent companies from using this technology.

Behringwerke must leap more hurdles, for instance in obtaining a licence to sell EPO, even if the production licence is granted.

But despite all the difficulties, Behringwerke is not considering pulling out of West Germany. "We're a Marburg company," said Faust. "We've invested DM10 million [\$6 million] in the plant already. It would be senseless if we could not use it."

Steven Dickman

Radiotelescopes in the United States, Europe, Australia and Japan will participate in the programme. Discussions are under way with the Soviet Union according to Haruto Hirose, professor of remote-sensing engineering at ISAS. The structure of the central cores of quasars, active galactic nuclei and other compact radio sources will be studied and their relative positions and motions obtained with "unsurpassed accuracy" ISAS scientists say.

Under an earlier plan, the antenna of the satellite would have been limited to a diameter of 5 metres because of the small lift capacity of ISAS's biggest rocket, the MU-3SII. But the Space Activities Commission in June approved a proposal by ISAS to develop a solid-fuel rocket with three times the power of the MU-3SII which has allowed ISAS scientists to double the planned size of the antenna to 10 metres, says Hirose.

Various ingenious designs for the antenna were presented by companies at the symposium last month, including an inflatable antenna design put forward by Kawasaki Heavy Industries and a mesh antenna design by Toshiba. But ISAS has decided to give the contract to Mitsubishi Electric Corporation for a tension truss design. NEC will handle the rest of the satellite and communications.

The total cost of the satellite is not yet certain but, according to Jun Nishimura, director general of ISAS, it will not be much more than that of previous ISAS satellites. They typically cost a few tens of millions of dollars, while the larger rocket for the launch will cost about ¥5,000 million (\$35 million).

David Swinbanks

Keeping the South Pole clean

Washington

A SERIES of accidents in the Antarctic this year, coming on top of accusations that US research bases are polluting the Antarctic environment, is causing increasing concern over US Antarctic policy. At a Senate hearing last week to assess the progress of the National Science Foundation's (NSF) clean-up operation, the debate widened into general discussion of how to protect the Antarctic environment. Next month, the Senate must decide whether to ratify the Convention on the Regulation of Antarctic Mineral Resource Activities (CRAMRA), which would in principle allow commercial mineral exploration and extraction to begin.

Even with less than a thousand researchers and a few thousand tourists in the area, damaging accidents have occurred. Peter E. Wilkniss, director of Polar Programs at NSF, told the Senate Committee on Commerce, Science and Transportation that the Argentinian vessel wrecked in January off Palmer station (*Nature* 337, 495; 1989) still contains 60,000 of diesel oil, which is slowly leaking into the ocean. James N. Barnes of the Antarctica Project said that two other accidents took place in February, involving the Peruvian ship *BIC Humboldt* and the British resupply vessel *HMS Endurance*. Although both accidents went largely unreported, either could have caused serious marine pollution.

The diesel spill occurred in a dangerous offshore channel near Palmer station despite warnings from US personnel. According to Barnes, Antarctic navigation is made more difficult by the lack of a single set of accepted charts.

The frequency of accidents and lack of international cooperation increases concern over what might happen if Antarctica is opened up to commercial mineral exploration. At present even prospecting is banned, although seismic surveys have CERN

Finland to join

London

FINLAND is to start negotiations with the European Laboratory for High-Energy Physics (CERN) with the aim of becoming the fifteenth member of the organization by 1991, it was announced in Helsinki on 29 August. Finland expects there to be a transition period after full membership during which its membership fee will gradually increase to its full contribution, calculated at almost 2 per cent of the CERN budget on the basis of spending figures for 1989. CERN welcomes the development, saying that it completes the geographical participation of Western European countries.

Ben Webb

been conducted for 'research' purposes. The ban was agreed as a temporary measure while CRAMRA was being negotiated. If ratified by 16 of the 22 member states, CRAMRA would allow mineral exploration and development, although each project would have to be approved by all members and strict environmental standards would be imposed.

Both Australia and France have expressed opposition to the convention, and a number of other Antarctic Treaty member nations are thought to be ambivalent. Australia favours a permanent ban on mineral development, but R. Tucker Scully, director of Marine Science and Polar Affairs at the State Department, said this would be not only undesirable but also politically impossible, because a consensus of nations would be required.

Most independent environmental organizations oppose CRAMRA; Barnes argued that CRAMRA's mere existence would spur commercial competition and accelerate development. But the Ocean Resources Institute takes a different view. Its director, Lee Kimball, feels that "international laws are more durable than political bans" and sees the regulations of the convention as necessary protection in the unwelcome but not unlikely event that mineral prospecting will take place.

Of more immediate concern to both environmental groups and the US government is the mess created by 30 years of research activities. Reports produced last summer by environmental groups (*Nature* 334, 643; 1988 and 337, 399; 1989) have accused NSF of polluting the unspoiled polar environment, and NSF has admitted that corrective action is needed. One problem is not knowing what has been dumped or released into the environment in the past: Wilkniss says that divers have stumbled upon bulldozers of which there is no record. Some progress has been made in cleaning up old buildings, outdoor storage areas and re-landscaping landfill dumps, but open burning of rubbish continues in contravention of applicable international law and raw sewage is still discharged into the sea. The NSF is requesting \$10 million in 1990 as the first increment in a \$180-million, five-year environmental health and safety initiative for its Antarctic Program.

Environmental issues will be high on the agenda when Antarctic Treaty states meet in Paris next month. Australia has proposed a "comprehensive environmental protection" convention, and the United States has suggested a more permanent infra-structure, in the form of a secretariat that could take responsibility for ensuring compliance with the regulations that exist.

Penelope Austin

ESA tries to change NASA's mind

Paris

SENIOR representatives of the European Space Agency (ESA) are this week returning to Washington in a further effort to persuade the US National Aeronautics and Space Administration (NASA) to abandon proposals to cut back on its Freedom space station (see *Nature* 341, 3; 7 September 1989). The plans, which could involve both an extension of the construction timetable and reductions in engineering specifications for Freedom, are unacceptable to both Japan and ESA and discussions in Washington last week failed to produce a solution.

Dr Frederick Engström, ESA's director of space station and platforms, does not rule out the possibility of changing construction schedules and is optimistic that an "equitable" agreement can be reached "if everybody is willing". But, he says, "the situation is much deeper than this".

The current proposals, which include a halving of available power, would, he says, mean that "there are really no resources available to operate anything but a US lab" and fears that there would be "no technical way to get these resources back within an acceptable time". Rescheduling, he says, would make the Columbus attached laboratory and free flying laboratory "like an appendix" to the main schedule and, from an engineering point of view, "unwelcome".

Among alternative cost-cutting proposals being tabled by ESA are the possibility that, if only one attached laboratory is feasible, this should be built by the partners (ESA and Japan) and not NASA. Also, the US flight tele-robotic servicer could be deferred and its functions performed by the Canadian-built robotic arm until the space station reaches its original specification.

Common mechanisms for decision-making exist between ESA and NASA and, says Engström, decisions should be based on consensus. But NASA does have the ability to make decisions unilaterally if no consensus is reached. Memoranda of understanding (MOU) signed between ESA and NASA should have ruled out the possibility of the proposed change in schedule and specifications. "We will have to look at these MOUs", says Engström, "but above all, we have an international governmental agreement which has been signed and ratified. We will have to see whether these agreements have to be adjusted or not."

NASA is expected to reach a decision by the end of this month. Meanwhile, ESA legal experts are looking at the implications of changes in the agreements. The new proposals will be high on the agenda at ESA's science programme committee meeting on 20 September.

Peter Coles

Who's hiding primary data?

Washington

THE International Union of Crystallography (IUC) this month published a policy statement urging scientific journals to ensure that when researchers publish analyses of proteins, nucleic-acid and virus structures, they simultaneously make the primary data available, to allow independent verification of the results.

Increasing delays in the disclosure of primary data are causing heated debate in the crystallographic community and giving rise to accusations ranging from sheer sloth to unjustified suppression of data obtained through publicly funded research.

The standard way to provide the primary data is through the deposition of the structure's atomic coordinates in the protein databank at Brookhaven National Laboratory. Many researchers believe that journals have a role to play in encouraging faster deposition of coordinates.

Last year, Frank Richards of Yale University wrote to several journals on behalf of almost 200 crystallographers urging them to "adopt and enforce" rules to make data more easily available. This would "ensure the preservation of important and expensively determined data which otherwise are likely to be lost". To

satisfy researchers reluctant to release the data immediately, the databank could be instructed not to release the information until a specified date.

But the issue divides journals. The editor of *Science*, Daniel E. Koshland, Jr, is willing to agree to these guidelines, but advocates a coordinated effort by all the relevant journals. The *Journal of Biological Chemistry* has strengthened its guide to authors to encourage deposition strongly. Others believe journals cannot act as 'watchdogs'. *Nature's* line is that there should be as few obstacles as possible between an author and publication. Some journals argue that granting agencies, for example, could require appropriate disclosure of data.

Even if journals agreed to require deposition of data in databanks, whether this should be done at the same time as publication of a paper is controversial. Some researchers argue that it takes several

years of refinement before a structure is ready to be entered into a databank: Ken Holmes, of the Max-Planck Institute for Medicine, says it is "general practice to withhold coordinates . . . and up to a point this is legitimate".

But Tom Steitz, of Yale University, sides with those who believe that if a structure is ready for publication, there must be data which can be released.

Steitz says the difficulty will be in finding some middle ground between those who call for immediate deposition of all data, even those not included in publications, and those who advocate a five- to ten-year time lapse while the structure is defined.

Rumours are now circulating in the crystallography community that researchers are selling coordinates to private companies instead of making them publicly available. Graham Derby of Wellcome says the debate is fuelled by a changing atmosphere in research, in which academics "are being encouraged to profit from anything that is potentially valuable".

Christine McGourty

BRITISH COUNTRYSIDE

Row ahead on conservation

London

A row is brewing between the British government and conservation organizations over plans to split the the Nature Conservancy Council (NCC) across the national boundaries of England, Scotland and Wales. The proposals first surfaced in the summer, when Nicholas Ridley, then Secretary of State for the Environment, said the proposals were designed to "improve the way we handle conservation of wild fauna and flora and the countryside".

Ridley said that the "circumstances and needs" of England, Scotland and Wales are very different and that the present system of nature conservation are inadequate. The responsibilities of the NCC, which looks after nature conservation throughout Great Britain, and the two Countryside Commissions (CC), one covering England and Wales, the other Scotland, would be altered. The CCs are responsible for enhancing the natural beauty of the countryside.

The functions of the three bodies will be "exercised in future by separate bodies for England, Scotland and Wales", Ridley said. "The conservation and countryside functions should be the responsibility of a single body in each country." But the government intends that because of the density of population and the greater pressure on land in England, the NCC and CC there will be retained as separate bodies. The changes "require legislation which we shall bring forward at the earliest opportunity", Ridley added.

One of the more contentious issues will

be the future of the scheme for designating and conserving "sites of special scientific interest", at present the responsibility of NCC; these range from habitats for valued animal and plants species to sites of geological or archaeological interest. Among other means of conservation, NCC administers a scheme for paying compensation to landowners for restraint in land exploitation, but more often the scheme has been resented (and even resisted) by those seeking to develop land.

The forthcoming debate will be an occasion when the British government will have further to define its plans for encouraging farmers to pay more attention to environmental considerations. But it is unlikely, at this late stage, that even those opposing the government's proposed reorganization will seek to put back the clock to when NCC had a research capability of its own, long since transferred to the Natural Environment Research Council.

But Sir William Wilkinson, chairman of NCC, says that the plans could "damage" the cause of conservation. He said conservation in Britain faces "complex challenges" and "mounting financial restraints" and that the proposals do nothing to help. "Wildlife issues are no respecter of geographical or political boundaries", he said, while the separation would "undermine" the scientific capability of the NCC and "prevent effective policy advice being provided to government". In their present form the proposals would "weaken" the NCC's capacity to play a "co-ordinated and positive role".

Ben Webb

GERMAN UNIVERSITIES

Green light for plan for young researchers

Munich

A PLAN to keep talented young researchers at universities received preliminary approval in West Germany last week.

Representatives of federal and *Länder* governments gave tentative approval to the plan, which would require an investment of DM6,000 million (about \$3,060 million) over ten years.

The plan, announced by Education Minister Jürgen Möllemann in May, is meant to prepare for an expected wave of retirements of university professors in the years 1995 to 2005 (see *Nature* 340, 252; 1989). Möllemann also plans to make funds specifically available to women in universities, where they have traditionally been under-represented.

As many as 5,000 young people would be supported in the programme's first year, and up to 50,000 would be supported at the end of the plan, said an Education Ministry spokesman.

Final approval will have to wait. A meeting originally scheduled for 21 September between Chancellor Kohl and the heads of the *Länder* at which the plan was to be discussed has been postponed because of Kohl's poor health. Steven Dickman

Considered view of Neptune

Washington

THE spectacular shot, shown below, of Neptune's ring system was taken by Voyager 2 as it looked back towards the planet. In sunlight reflected at a glancing angle, the fainter rings and dust bands are clearly visible, but a comparison with pictures of the rings taken as Voyager approached Neptune (see *Nature* of 31 August) shows an intriguing difference: the outermost ring, significantly brighter on the approach shots, shows up here a little fainter than the prominent inner ring.

The difference is a clue to the composition of the rings. Particles bigger than the wavelength of light, about half a micrometre, scatter most of the radiation incident upon them directly backwards, towards the light source; smaller particles scatter more light forwards than backwards. Thus most of the ring and disk material around Neptune, except for the outer ring, is 'dust' — although a more apt comparison for the size of the particles would be 'smoke'.

An extended disk of fine material was presaged by another Voyager experiment. The plasma-wave detector, two long antennas arranged in a V, also acts as a dust detector. As Voyager 2 speeds on at several miles a second, any dust particles it encounters vaporize on impact: the ionized shock front from this tiny explosion carries a characteristic burst of radio noise, which the plasma antenna picks up.

Because the frequency of the radio noise from the dust impacts is in the kilohertz range, it can be replayed directly as an audio signal: this led to a moment of

drama in the Neptune encounter. As Voyager 2 passed through the equatorial plane on its way to closest approach, at about twice the radial distance of the outer ring, the hissing and crashing of the dust was played live through loudspeakers at the Jet Propulsion Laboratory. As the noise reached a maximum, it abruptly ceased; Voyager's radio signal was lost behind Neptune, to be recovered 45 minutes later.

Calculations of the encounter rate with dust around Neptune's equator suggested a maximum density of 0.003 particles per cubic metre. This is more dust than found around Uranus, and its presence, along with the 'clumpy' outer ring, reinforces the idea that Neptune has had a more interesting recent history than its sister planet. Explaining the denser segments in the outer ring will be the hardest task: two theories already on the books both seem to be misses. In the first, two guiding satellites, one at the same orbital radius as the ring and one just inside, create perturbations to the gravitational field that allow material to reside stably in restricted arcs; at Neptune, none of the four new satellites near the ring is in the right place. In the second theory, a satellite with an orbit inclined to the ring plane creates similar perturbations, and 1989N6 is indeed in a circular orbit tilted at 5 degrees. But it seems too small for its gravitational influence to be able to stabilize the dense ring segments.

In another respect, Neptune and Uranus are twins: both have magnetic fields whose axis is tilted, at about 50 and 60 degrees respectively, markedly from the rotation axis. This presumably has something to do with the internal structure of the planet. Unlike Earth, which has a hot, semiliquid core whose internal motions generate the terrestrial magnetic field, the rocky cores of the giant planets are small, cold and inert. Outside this core there is probably a thick layer of liquid water existing, because of the huge pressure, in an ionized phase. The dynamo for these planets may reside in this turgid but mobile region.

The tilted magnetic field complicates Neptune's magnetosphere, and its interaction with the solar wind. Detectors on Voyager found evidence for charged particles trapped by the magnetic field, but the density was unexpectedly low. The magnetosphere, whose geometry is con-

trolled by a tilted magnetic pole rotating about a planetary axis itself tilted with respect to the plane of the Solar System, may not be permanently isolated from the solar wind: from time to time, particles trapped within it can be swept out into interplanetary space.

David Lindley



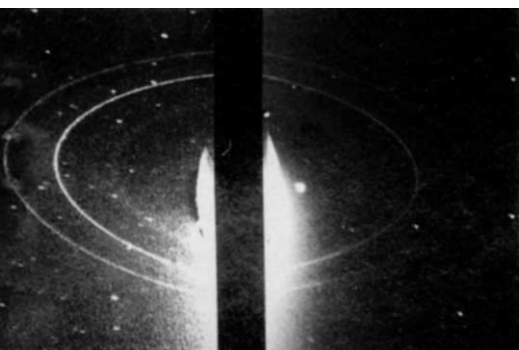
A false-colour composite of UV, violet and green reveals the altitude of Neptune's clouds. The bright pink cirrus clouds are above most of the atmosphere, which absorbs UV radiation and darkens the deeper clouds. The Great Dark Spot appears blue; it lies at an intermediate height from which blue light, but not UV, can escape.



A false-colour view of Triton, taken on Voyager's approach. The pink hue approximates to the real colour of methane ice, but the blue represents UV light: sadly, Triton is not a blue moon. The strip below, a composite of several shots, shows the variation in terrain. The feeble heat of Triton's summer is thought to sublimate surface ices, breaking up the smooth surface into the blotchy covering at left.



The largest of Neptune's 'new' moons, 1989N1, turned out to be oddly shaped. At about 250 km in radius, and thought to be made of deep-frozen water ice, it could not be any bigger without its own gravity forcing it into a spherical shape. Below, the rings in their full splendour, seen as Voyager looked backwards after passing by Neptune.



Towards the day of hard choices

Jon File, Annamia van den Heever and Stuart John Saunders

When apartheid breaks down (as it must), South Africa will need a new, non-racial educational system. Moves to that end are already afoot.

LAST week the National Party was returned to power in South Africa. Although the party has promised reform, most South Africans expect that the essentials of the system of apartheid that divides South Africa from the rest of the world, and its people from one another, will not be fundamentally changed.

Since 1953, when Hendrik Verwoerd first enshrined racially divided education in the Bantu Education Act, education policy has been the cornerstone of 'separate development'. Rejection of apartheid education sparked the Soweto student uprising of 1976 and the ensuing 13-year-old crisis that shows no signs of ending. For several months now many schools have been closed. Police are regularly in school playgrounds and some university campuses have at times been occupied by security forces. F. W. de Klerk (himself until this month Minister of National Education) and his incoming administration have inherited this legacy—for them, as for their predecessors, educational issues will continue to be central as apartheid continues to disintegrate in the face of internal and external pressure.

Background

The National Party's policy makers, who codified the principles of apartheid into legislation from 1948 onwards, were quite clear about their intentions. White, coloured, Indian and African* were to live apart, and whites were to retain political and economic dominance. In consequence, education should be to the appropriate level. Verwoerd was explicit that, for Africans, this meant no more than the minimum training required for manual labour. Whites, on the other hand, had to be educated to lead and, in the case of Afrikaners, to preserve the Afrikaner heritage. 'Christian National Education' was based on the view that Afrikaners were a distinct Christian nation, with a language, history and culture of their own. The language of apartheid may be more sophisticated now than it was 40 years ago, and some changes have occurred, but racially segregated education remains the foundation of South Africa's education structure.

* In this article we use the term 'black' to refer collectively to all those denied full political rights by the system of apartheid. The racial designations 'white', 'coloured', 'Indian' and 'African' are those used by the state in administering apartheid.

Under the philosophy of Christian National Education, the school system is governed by a multiplicity of education departments, each with its own examination system, syllabuses and standards. A common feature is authoritarian control. The same philosophy has also divided the

Table 1 Comparative statistics for South African schools, 1987

	White education	African education
Pupil:teacher ratio	16:1	41:1
Underqualified teachers (less than standard 10 plus a 3-year teacher's certificate)	2%	87%
Per capita expenditure	R 2,508	R 477
Standard 10 pass rate	94%	56%

R, South African Rand

Source: J Hofmeyr & R. Spence *Bridges to the Future* in *Optima* Vol. 37, No. 1 (1987).

universities. Although some tertiary institutions were founded specifically to service apartheid, the four English-speaking open universities (Cape Town, Natal, Rhodes and Witwatersrand), as well as the University of the Western Cape, have vigorously opposed the imposition of apartheid framework. The Extension of University Education Act (1959), which introduced a 'permit system' for black students wishing to study at universities designated 'white' by the government, was vigorously fought. Legislation in 1983, which replaced the notorious permit with a ministerially determined 'racial quota', was also resisted, with the result that although the provision is on the Statute Book no quota has ever been enforced.

These five universities have consistently opposed the encroachment on academic freedom and freedom in general: from the imposition of successive states of emergency, banning and detention without trial, censorship and restrictions of the press, to restrictions and bannings of individuals and organizations. In 1987 the Universities of Cape Town, Western Cape and Natal successfully challenged, in the Supreme Court, a government attempt to link state subsidy to the control of political activities on university campuses. The five universities have usually stood alone, although staff and students at other universities, particularly the predominantly black universities, have not been silent.

Those tertiary institutions that have made a principled stand against apartheid

remain inexorably bound to the wider crisis in education. Black students at the open universities are tremendously disadvantaged, and academic support programmes are essential. Black candidates for university admission often require full bursary support if they are not to be denied further education for purely economic reasons. Political conflicts often erupt in the pockets of relative freedom on the open campuses. For these and other reasons, educational issues in South Africa have to be seen on a broad canvas.

Schools

A recurrent feature of popular uprisings since 1976 has been school boycotts by black pupils protesting against the iniquities of apartheid, including the education they receive. Although the state response to such boycotts in 1976, 1981 and 1984–1985 was to increase further the already harsh restrictions, the government announced in 1986 its intention to achieve parity in education according to a Ten Year Plan. This would have increased expenditure on education by 4.1 per cent per annum in real terms. But this year the Minister of Education stated that this goal could not be achieved. "As far as education is concerned we are in a tight spot", he told parliament, citing the main reason as the low growth in the economy, due partly to sanctions and disinvestment (*Hansard*, No. 11, cols 5555–5559; 1989). Thus although there have been some improvements in recent years (particularly in the greatly increased financial provision for black education which has expressed itself in better qualifications and salaries of black teachers and facilities for black schools), huge disparities remain. Table 1 shows the stark differences between white and African schools in pupil:teacher ratios, teacher qualifications, per capita expenditure and pass rates. At the same time, the white school-going population will decline from 954,000 in 1987 to 899,000 in the year 2020, while projections for the African school-going population show that it could more than double from 6,645,000 in 1987 to 14,977,000 in 2020. (Figures from E. Dorstal's *Notebook on Educational Trends and Perspectives*, Institute of Futures Research, University of Stellenbosch; 1988.)

More than a decade of instability coupled with inadequate resources has led to a situation in urban black schools today in which very little education is taking place.

Drop-out rates are high. Teachers, caught between the administrators of apartheid and their own communities, are demoralized. Pupils lack the will to learn. Parents have little control and discipline has almost completely broken down. Illiteracy in South Africa is estimated to be as high as 50 per cent.

Universities

A sustained crisis in secondary education obviously has severe effects on tertiary education. These effects and the response to them vary, and are modulated by the place of individual institutions within the framework of apartheid education.

Some 420,000 students are currently enrolled at higher-education institutions: 300,000 at universities, 68,000 at technikons and 53,000 at teacher-training colleges. These figures are part of a trend of rapid increase: in 1966 70,000 students were registered in South Africa's universities, while over the past three years alone an additional 60,000 students have been awarded places. Over the next ten to fifteen years the expected three-fold increase in the number of matriculants, the great majority of whom will be African, will continue this exponential demand for university places. But rather than challenging the structure of separate education, the increases in enrolment are having little effect on the system of racial segregation and discrimination.

Despite the abolition three years ago of the permit system, which restricted access to universities on the basis of race, the effect of apartheid is still starkly evident (Table 2). Ninety-eight per cent of students at universities designated African in the original structure of apartheid education are African, and 96 per cent of students at Afrikaans or dual-medium universities, are white (the University of South Africa, or UNISA, is a non-residential, distance-learning university that has always been open). Looked at another way, although slightly more than a third of current university students are African, 51 per cent of them are at the eight African universities, 39 per cent of them are studying by correspondence through UNISA, 2 per cent of African students are at the University of Durban Westville (originally an 'Indian' university), a further 2 per cent are at the University of the Western Cape (intended as a 'coloured' university) and 5 per cent are at historically 'white' universities (almost all of them at the English-speaking universities). These figures must be set against South Africa's demographic structure: there are over 30 students per thousand of the white population and fewer than three students per thousand of the African population (see *Macroaspects of the University within the Context of Tertiary Education in the RSA*, Committee of University Principals; 1987). It is difficult to escape the conclu-

sion that, despite changes in legislation, Verwoerd's vision of separate education has been successfully realized in the university system as well as in the schools.

Additional biases are less immediately evident, but are of crucial importance. With the exception of Medunsa (a medical university established for blacks), the natural sciences are best developed at the 'white' universities. Together with the low standards of mathematics and science education in African schools, this has meant that Africans have mostly studied the humanities, social sciences and education and, to a lesser extent, law and commerce. Similarly, postgraduate study is concentrated at the 'white' universities: only 5 per cent of South African masters and doctoral students are African. It is not surprising that 91 per cent of the permanently appointed academic staff in South Africa in 1986 (the most recent year for which comparative figures are available) were white and less than 5 per cent African.

As with the schools, the present basis of financing is not sufficient to allow universities to do much about these inequalities. While the South African government allocates a respectable 20 per cent of public expenditure to education, and universities receive a sizeable 15 per cent of this, the joint consequences of past neglect and future demand have created a financial crisis. Thus the government has cut back its intended support for universities by some 25 per cent. Per capita income to universities is increasing at about half the rate of inflation. Government budgetary planning is based, at most, on annual increases of 1–2 per cent in real terms. Government policy is to rationalize the university system and to limit growth through the exclusion of marginally qualified entrants.

Although these forces act in concert upon the university system their effects vary from institution to institution. The educational challenges arising from inadequate schooling are faced primarily by the black universities, and on a lesser but perhaps more complex scale by the four English-speaking open universities given the wide range of educational backgrounds of their first-year students. Acute demand for university places is faced only by those universities committed to improving access for African students. The financial situation is worsened for these same universities. South Africa has no national student financial aid programme. The costs of a year of study at a residential metropolitan university exceed even the minimum annual wage levels negotiated by strong Trade Unions with more enlightened employers. In this context, universities committed to increasing African access have had to invest heavily in financial aid programmes, and to give them precious fund-raising priority.

Current debates

The crisis in South African education has aptly been described by Hartshorne as an interregnum in which (following Gramsci) "the old authority is dying and the new cannot be born" and where, as a result, "a great variety of morbid symptoms appear" (*Post-Apartheid Education*, to be included in *Critical Choices for South African Society*, Oxford University Press; 1989). The South African government has consistently repressed organizations working towards a non-racial future. Hundreds of people involved in education have in recent years been detained without trial or have had severe restrictions imposed upon them. Nevertheless, there has been a great deal of debate about alternatives to apartheid education.

At school level the debate has centred on fundamental issues. Among them are the extent to which education should serve the individual, society and the state, the nature of a 'relevant' curriculum and who should determine relevance, and the question (given that South Africa is multi-lingual) of whether students should receive mother-tongue education and for how long. When in a school career should a common language (which one?) be used as the medium of instruction? Given the demographic realities of South Africa, is mass formal education possible? Who should pay for it — the state, private enterprise or both? Where does authority end and authoritarianism begin?

Both Christian National Education and Western-style liberal education are under scrutiny. Liberal education is viewed by some as fostering elitism and competitiveness, and entrenching capitalism, and as a vestige of colonialism. An emerging, alternative philosophy is 'People's Education', which can be regarded as a set of proposals which flesh out the educational principles contained in the Freedom Charter (a declaration of political aspirations and intent formulated during the Congress of the People, organized by the African National Congress at Kliptown in 1955). The structures envisaged in People's Education are unitary and non-racial, and organized so that all parties involved in education — students, parents, teachers, workers — have a part in its management.

These debates are echoed at the university level. Questions are being raised about how the curriculum and research priorities should reflect South Africa as a part of Africa; the role of the curriculum and the university in reproducing inequalities, not just by race, but also by gender and class; the balance between research generated by disciplinary interests and that initiated in response to national needs; the extent to which teaching and research should concentrate on the needs of government and big business at the expense of those of the broader community; links between universities and the

military and the armaments industry; and the democratization of university power structures and many other issues.

A wide range of educational programmes has developed in parallel with this broad debate. The English-speaking open universities and the University of the Western Cape are islands in a hostile environment, but it has still been possible to initiate imaginative projects in educational development, academic support for educationally disadvantaged students, and devising appropriate selection criteria. Large investments in personnel and in time have been made in order to build a non-racial ethos and appropriate industrial-relations practices. Financial-aid programmes have been introduced. Student residences were fully integrated in spite of the Group Areas Act.

There are also encouraging developments outside the universities. Within the state school system there have been calls for open schools from white-teacher organizations and white parents, and there are moves towards teacher unity on a non-racial basis. Some white-parent organizations, and business, have been questioning the relevance of the education white pupils are receiving. Innovative programmes are being negotiated and adopted elsewhere, in private non-racial schools, in organizations involved in informal, adult and worker education, in literacy programmes, in upgrading programmes for teachers, to name but a few. Attention is also being focused on the education departments in the non-independent 'homelands' and 'independent' states where 70 per cent of African pupils are at school. It is now recognized that the 'homelands' cannot be wished away and that their peripheral nature to the central system could offer opportunities.

Characteristic of the innovations taking place are painstaking attempts at community consultation in order to legitimize not only the programmes but the process used to derive them.

Underlying all of this activity is the explicit recognition that political transformation is a pre-condition for any real educational solution. The government's Ten Year Plan for education, its funding formula for universities and the noticeable lack of effect that changes in legislation have had on enrolment patterns across the university system all show that apartheid education cannot be reformed. Any real solution must be a post-apartheid solution.

Along with this political transformation will come hard policy choices. Should increased education expenditure take priority over programmes addressing health care, agricultural development, housing, water and sanitation, and energy and electrification? Post-apartheid South Africa, even with an improved economy and savings from reduced defence expenditure and the scrapping of separate sys-

Intended population group in terms of government policy	Number of universities	Total enrolment April 1989 (nearest 1,000)	Enrolment by population group (%)			
			White	Coloured	Indian	African
White						
(a) English medium	4*	47,000	74	5	9	11
(b) Afrikaans or dual medium	6†	67,000	96	2	<1	1
Coloured	1‡	12,000	2	77	4	18
Indian	1§	7,000	4	2	61	33
African	8	54,000	<1	<1	<1	98
UNISA	1¶	113,000	50	5	9	37
Total	21	300,000	52	6	6	35

* Cape Town, Natal, Witwatersrand and Rhodes. † Orange Free State, Port Elizabeth, Potchefstroom, Pretoria, Rand Afrikaans and Stellenbosch. ‡ Western Cape (UWC). § Durban Westville (UDW). || Medical University of South Africa (Medunsa), North, Zululand, Bophuthatswana**, Transkei**, Venda**, Fort Hare**, Vista. (** 1988 figures are used. These universities are situated in 'independent states'.) ¶ UNISA, although administered by the white education department, has traditionally been open to all. Source: Department of National Education.

In 1986 the South African Institute of Race Relations estimated the population of South Africa, based on the 1985 census but adjusted for undercount, to be: African 24,901,139 (74.1%); Indian 878,300 (2%); coloured 2,881,362 (8.6%); white 4,961,062 (14.7%).

tems of administration, is unlikely to be able to afford a university system that allows access at the rates currently enjoyed by white students, or a school system funded at the same level as white schools. What relative priority should literacy and numeracy programmes, pre-primary, primary and secondary education, adult and worker education, technikons and universities receive within the education budget? The issues are complex, the answers as yet unknown. What is clear, however, is that difficult choices are inevitable, and that the process by which such choices are made will need wide support and legitimacy if there is to be success.

Such support and legitimacy will be facilitated if those sectors of South African society that are currently excluded from policy debates and decision-making structures are drawn into them. Although the tempo of immediate political struggle often lowers the priority given to future-orientated projects, their importance needs to be emphasized. All concerned with education have to begin to address the questions of how the universities will deal with national personpower, with research and development needs, with ensuring an indigenous scientific technological and professional base, with contributing to the formation of a new South African nation, and with redressing the historical inequities of access across all universities and at all academic levels.

Competing social priorities in post-apartheid South Africa will probably mean that growth in university places is restricted, or even reduced. In higher education, preference may need to be given to the expansion of technikons, and to the natural sciences in both technikons and universities. Given that full-time student enrolment may be limited, increasing emphasis on distance education, part-time study, continuing and adult

education, and increased access to universities through the wide dissemination of research results and a broader availability of university expertise through consultancy and community projects may all become priorities. These issues need careful research and discussion before the day of hard choices dawns.

Conclusion

Although the government received a white mandate for reform in last week's election, the struggle for a non-racial, democratic South Africa will be a long one. The "old authority is dying" and there are indeed many "morbid symptoms", but there are also the signs of new life, a common culture on which a strong education system can be built.

Those within the South African educational system who have grasped the opportunity to begin to anticipate the future need help. Among the institutions and sectors involved are emerging non-racial schools, future-directed literacy and education programmes, the English-speaking open universities (of Cape Town, Natal, Rhodes and Witwatersrand) and the University of the Western Cape, and staff groups at the black universities who take the brunt of the challenge and work with most of South Africa's African students. Financial aid for black students is a key priority if lack of money is not to become the main inhibitor of black access to university. Carefully selected support for these and other areas will hasten the day when a unitary non-racial education system, that serves the needs of the new South African nation, will be achieved. □
Jon File, Annamia van den Heever and Stuart John Saunders are at the University of Cape Town, Rondebosch 7700, South Africa.

ACKNOWLEDGEMENTS. We thank colleagues involved in South African education for helpful discussions, and particularly Martin Hall for his editorial advice.

Policing animal experiments

SIR—A letter from Mr Clive Hollands appeared in *Nature* (339, 248; 1989) stating that experiments published in this journal earlier in the year (337, 265; 1989) would have been illegal had they been carried out in the United Kingdom. It took two weeks for *Nature* to point out that this was an expression of opinion and not legal fact (339, 407; 1989). Your error in publishing the letter without proper consideration was compounded by the inappropriate and emotive heading for your leading article ("Beastly experiments") the following week. This headline puts *Nature* in the same category as *The Sun* and *Sunday Mirror*. The antivivisection campaign of the *Sunday Mirror* against Professor Colin Blakemore was recently investigated by the Press Council, which found that the basis of the campaign was false and required the publication of that adjudication.

Unfortunately, much of the subsequent correspondence (for example *Nature* 340, 180; 1989) seems to have taken more notice of Hollands' incorrect assertion than of your belated article setting out the correct position. It is important that your readers, whatever their personal opinions about the use of animals for research, should know that in the United Kingdom there is a new Animals (Scientific Procedures) Act (1986) which controls the use of such animals. All scientists in the United Kingdom using animals for research must hold a personal licence and be covered by a separate project licence, both issued by the Home Secretary after appropriate application and consideration. Section 5(4) of the Animals (Scientific Procedures) Act (1986) states: "In determining whether and on what terms to grant a project licence the Secretary of State shall weigh the likely adverse effects on the animals concerned against the benefit likely to accrue as a result of the programme to be specified in the licence". The experiments Hollands complains of have not been put to the Home Secretary and therefore it is not known whether they would be allowed in the United Kingdom. Hollands is a member of the Animal Procedures Committee that advises the Home Secretary on matters relating to the licensing of scientists to carry out animal experiments. He therefore knows or should know what the law states. There is good reason to believe that Hollands was advised before sending his letter to *Nature* that it would be inappropriate for him to claim that these experiments would be illegal in the United Kingdom.

Nature has done a great disservice to biomedical research in this country. You should attempt to repair the damage by a proper apology to the French scientists whose work you have effectively repudi-

ated and by a discussion of the likely adverse effects on both medical and fundamental research of your too ready acceptance of an antivivisectionist's minority viewpoint. Blakemore has been and still is the subject of an aggressive personal attack by antivivisectionists. This attack now appears to be spreading to his colleagues around the world. Their work on the development of the visual system is of the highest order and has led to important advances in the understanding of both basic mechanisms and clinical problems that may arise in the young. It is high time that Blakemore and those other scientists, who have been pilloried for their work by a tiny but aggressive minority, were properly and publicly supported by international organs of communication such as *Nature*, and by government and charitable organizations that provide much of the financial support for research involving animals.

NORMAN R. SAUNDERS

Department of Physiology &
Pharmacology,
University of Southampton,
Southampton SO9 3TU, UK

Reprints' function

SIR—In spite of the plethora of letters to *Nature* regarding the pros and cons of honouring reprint requests, a major point seems to have been missed by those against. Gone are the days when a scientist could make a leisurely trip to the library, leaf through the three or four journals of interest to him/her, and photocopy the few articles of interest. There has been a tremendous 'information explosion' in both the number of journals in any particular field and the number of articles. To a scientist who wishes to maintain any breadth of knowledge, even within a highly restricted field, this is simply no longer possible. It would require two or three days each week of exclusive library work.

That increasing numbers of scientists are making use of commercial computer searches and associated reprint requests is a reflection of this change within the scientific community. It is neither a reflection of laziness nor of selfishness on the part of the reprint requesters, as suggested by Peter Johnson (*Nature* 339, 170; 1989). Rather, it is a reflection of time economy, combined with a desire to keep up with the literature.

On the other hand, some scientists, according to Johnson, apparently resent the "throwing [of] the burden of time and cost onto the author". The time taken to mail reprints is considerably less than in searching out an article and mailing it, especially as most requests are accom-

panied by easily used return addresses. Thus the real problem appears to be the cost. This seems to me to be a predictable cost that granting agencies are willing to accept as an appropriate item on a budget. For those without grants, it could be an item for negotiation with their departments as an administrative cost as essential to any science department as writing paper or secretarial support.

Scientific endeavours are meaningless if not communicated to others in the appropriate scientific community. Given the nature of contemporary science, the problem of this communication (of which the honouring of reprint requests is an important element) is not a trivial issue. Perhaps it is time for the recognition of the necessity of the dissemination of scientific information, and a general agreement as to whose shoulders the responsibility for this dissemination should fall upon.

MATHEW T. MARTIN-IVERSON

Department of Psychiatry,
University of Alberta,
1E7 44 W Mackenzie Health Sciences
Centre,
Edmonton, Canada T6G 2B7

SIR—On the question of reprint requests, many Russian scientific journals are translated into English from cover to cover, but authors usually have no free reprints of the translated papers and so cannot usually satisfy reprint requests. The authors can only send reprints of the Russian originals.

BORIS V. SHESTOPALOV

Institute of Cytology,
Academy of Sciences of the USSR,
Leningrad 194064, USSR

No to South Africa

SIR—I am one of those who have stuck their necks out and said "No" to science from South Africa. I have done it because I believe that the inviolateness of the practice of science is not absolute; its precedence is relative and must be judged against the enormity of the racial injustice in that country. Racial injustice is being practised in many countries, but it is the scale of it in South Africa that stands out. One of the arguments against boycotting South African science is that it harms innocent scientists. I would weigh this against the harm being suffered by millions of innocent blacks at the hands of a much smaller number of whites. If the black leaders believe that sanctions and boycotts will help to change their situation for the better, then I will support them in this way. Who else should we listen to but the victims?

JOHN ORMEROD

Department of Biology,
University of Oslo,
PO Box 1050, Blindern,
N-0316 Oslo 3, Norway

Does the Earth really move?

SIR—John Durant (Britain's first Professor of the Public Understanding of Science) *et al.* (*Nature* **340**, 11–14; 1989) complain that “only 34 per cent of Britons and 45 per cent of Americans ... were able to state correctly both that the Earth goes round the Sun and that it takes a year to do so”. They clearly implied that the 30 per cent of Britons who responded that the Sun goes round the Earth are mistaken, and they actually stated that “most of the public appear not to have caught up with Nicholas Copernicus and Galileo Galilei”. But look what Albert Einstein and Leopold Infeld wrote in their book *The Evolution of Physics* (Cambridge University Press, 1938, p. 224):

Can we formulate physical laws so that they are valid for all coordinate systems (cs), not only those moving uniformly, but also those moving quite arbitrarily, relative to each other? If this can be done, our difficulties will be over. We shall then be able to apply the laws of nature to any cs. The struggle, so violent in the early days of science, between the views of Ptolemy and Copernicus would then be quite meaningless. Either cs could be used with equal justification. The two sentences, “the sun is at rest and the earth moves”, or “the sun moves and the earth is at rest”, would simply mean two different conventions concerning two different cs. Could we build a real relativistic physics valid in all cs; a physics in which there would be no place for absolute, but only for relative motion? This is indeed possible!

This important lengthy passage is quoted in full lest the incredulous reader might not trust this writer's word. Most of the public may not have caught up with Copernicus and Galilei, but it appears that most of the professors have not caught up with Einstein. Should not the public expect professors of the public understanding of science to be acquainted with the most basic facts of science before they complain that most of the public are ignorant of scientific matters?

THEO THEOCHARIS

Department of Physics,
Imperial College,
Blackett Laboratory,
Prince Consort Road,
London SW7 2BZ, UK

Reasonable doubt

SIR—Eric S. Lander's Commentary and your accompanying leading article (*Nature* **339**, 491 and 501; 1989) raise the question of reconciling scientific and legal thinking. However, while the mere fact of the appearance of these publications demonstrates that science is pursuing its usual practice of self-criticism, I have seen no evidence that the law is doing anything

to modify its practices in the light of scientific advance.

I wish to raise here two issues involving probability, which figures in both the contributions referred to. While most would agree that someone cannot be “60 per cent guilty”, it is quite possible that there may be an x per cent probability of guilt. In some cases the probability can be quantified with some exactness, and the question then arises — what probability represents the ‘reasonable doubt’ beloved of the legal profession?

In my experience, lawyers do not want to address this question, perhaps because it takes the matter out of their hands and puts it in the hands of scientists. But surely a legal consensus should exist, and be known to all members of the profession.

A second issue (which crops up quite frequently in the courts) is where death rates due to a given illness in a particular cohort are raised above the normal level by local external circumstances. When the arguments are purely statistical, we may know that $n+m$ individuals have died instead of the expected n , but it is impossible to determine whether any particular individual is one of the m or one of the n . In other words, it is impossible to say in any individual case whether the external circumstances were or were not the cause of death. I have never heard of this argument being raised in any of the large number of cases involving this type of circumstances that have been or are being tried.

It will be interesting to see whether lawyers are prepared to exercise the same sort of self-criticism as do scientists, and review their attitudes to circumstances such as those described above (there are many others) where science is at present either ignored or distorted.

GEORGE F. WOOD

Parkfield, Charlton, Tetbury,
Gloucestershire GL8 8TS, UK

Complications

SIR—In a News and Views article (*Nature* **338**, 457; 1989), John Maddox uses the word “complicated”, but I think “complex” is the right word. The word ‘complex’ is now a technical term used by physicists and computer scientists. A non-living static structure can be very complicated, but it need not be complex. Complexity results from dynamics. If a structure can, in principle, be described completely, it is not complex. Complicated non-living static structures are not complex because they can be described with sufficient accuracy in a finite time.

Complexity of dynamic systems is connected with unpredictability. In classical mechanics, it is assumed that the entire future is predictable in a deterministic system. But this is true only for linear deterministic systems.

Linear systems are those that can be divided into parts, and each part can be studied in isolation. Not much information is lost when the system is divided. The Solar System can be linearized in the sense that Earth–Sun, Earth–Moon, Sun–Jupiter subsystems of the Solar System can be studied in isolation without significant loss in the accuracy of the results. That the eclipses can be predicted so accurately exemplifies that linear systems are not complex.

Nonlinear deterministic systems, on the other hand, are complex. “Deterministic chaos” (Gleick, J. *Chaos*, Heinemann, London, 1988) is an example of a complex system. A chaotic system can be modelled with simplistic deterministic equations, yet the model cannot be used to predict the long-term time evolution of the system. It can be proved rigorously that the equations of the model can predict the entire future of the system exactly and uniquely *if we can specify the initial conditions exactly*. In practice, however, we can never determine the initial conditions exactly in real systems.

The crucial difference between chaotic and ordinary dynamic evolution is the effect initial error has on predictability. In ordinary evolution, errors grow linearly and they can be corrected by subsequent observation. In chaotic evolution, errors grow exponentially rapidly and cannot be corrected. Chaotic systems are thus complex. This is so in spite of the fact that they can be specified by simple equations.

SAIFUL ISLAM

Research Unit Gottstein,
Max Planck Society,
Frankfurter Ring 243,
D 8000 Munich 40, FRG

Before the event

SIR—It seems that you are caught on the horns of a philosophical dilemma (“Down with the Big Bang”, *Nature* **340**, 425; 1989). You reject an oscillatory view of the Universe at least partly on the grounds that “the ultimate origin of our world cannot be discussed”. You therefore reject as unsatisfactory an endless sequence of causes in a time which has no beginning. In the preceding paragraph, you object to the Big Bang because it implies the existence of an instant before which there was no time, and which is not susceptible to discussion of causes preceding the event. Surely this objection applies to any discussion that presumes to account for the “ultimate origin of our world”?

I have always thought it unlikely that a defence of the First Cause would ever appear in your columns. Now, however, I am not so sure.

MICHAEL WEBB

Molecular Biology Laboratory,
Sandoz Institute for Medical Research,
5 Gower Place, London WC1E 6BN, UK

Racemization rationalized

Do chiral partner molecules crystallize together as racemic crystals or as separate enantiomers? There is no general rule, but now at last there are the beginnings of a working model.

ONE of the heroic true legends of our time is that Pasteur demonstrated that asymmetrical carbon-based molecules are optically active by sifting through the small crystals obtained from a solution of a racemic mixture, putting those with gross left-handed symmetry to one side and those with right-handed symmetry to the other. That was the first proof that molecules that are chiral (mirror images of each other at a centre of asymmetry) are physically distinct, but it also shows that in Pasteur's case, the enantiomers crystallize separately. That is not a general rule. Just as often, the enantiomers crystallize into a racemic crystal with equal numbers of asymmetrical molecules.

All that is, of course, the familiar stuff of the elementary textbooks, providing telling illustrations of the value of Gibbs' phase rule and of much else in physical chemistry; small proportions of one enantiomer will indeed depress the melting point of the other, for example. But how to calculate the behaviour of particular pairs of enantiomers? Will enantiomers preferentially crystallize separately or together, for example? The phase diagram for the first case is the simpler; whichever enantiomer is in excess will be the first to crystallize from the melt, driving the remaining liquid towards 50:50 composition. If, on the other hand, a racemic crystal can exist, racemic crystals will appear directly from cooling melt with a composition centred on equality, and there will be two eutectic points on either side of 50:50 composition at which the solid that appears will be a mixture of racemic crystals and those of the enantiomer. Can the details of this behaviour be predicted?

Plainly that is not a simple question. Something needs to be known of the energy of interaction between enantiomeric pairs in various conformations and then, usually still more difficult, of the way in which they will be ordered in a crystal. Now David Andelman, of Tel Aviv University and the Collège de France, has made a start on the problem by constructing a model of how enantiomeric pairs may behave in two dimensions (*J. Am. Chem. Soc.* **111**, 6536; 1989).

Andelman explains that the two-dimensional version of the problem is not without interest in its own right. Membranes in biological systems are, for example, loaded with chiral phospholipids, while the inter-molecular interactions controll-

ing the properties of, say, a two-dimensional Langmuir-Blodgett film will at least be a sub-set of those effective in three dimensions. Indeed, two-dimensional studies should make it possible to define the interaction energy as a function of intermolecular distance. Andelman offers a model which is a neat way of ordering ideas likely also to stimulate further work.

Andelman uses the simplest possible model of a chiral molecule, centred on a tetrahedral carbon atom linked to four distinct groups, of which three are supposed to be anchored by their chemical affinity to a two-dimensional layer (think of it as a membrane) while the third (visualized as, say, a long aliphatic tail) sticks out perpendicular to the surface. If the three groups bound to the layer are A, B and C, one enantiomer is represented by molecules in which A, B and C appear in alphabetical order reading, say, clockwise around the triangle; the other is the set of molecules in which the order appears as A, C and then B. After suitable rotation of this 'tripod' molecule, the 'foot' of one molecule in the layer is a mirror image of the other.

It is not disparaging of Andelman's work to say that he has constructed what must be the simplest model there could be of an aggregation of such simple molecules: even that is difficult enough. The triangular feet of the molecules are supposed to interact only when two triangular faces are opposed to each other, which implies that only three orientations of the feet, differing by $2\pi/3$, are allowed. This has the consequence that the symmetry of any two-dimensional crystal that may be formed is hexagonal. The next step in the simplification process is to allow only for the interaction between groups in the triangular feet which are directly opposed to each other; interactions between next-nearest and more distant pairs of neighbours are neglected. Then the trick is to construct a partition function and derive the thermodynamic properties of the system from that.

Remarkably, some provoking generalizations (for the two-dimensional case) nevertheless appear to be possible. The parameters of the problem are the energies of interaction between the nine possible pairs of closely opposed groups, of which only six are independent (BC equals CB, for example). Andelman finds that, if all these interactions are due to

Van der Waals forces, chiral segregation is energetically less favoured than the formation of the two-dimensional versions of racemic crystals. Or, at least, that is how it seems. In this case, the six independent parameters are reducible to three because the pairwise interactions may be expressed in terms of the polarizability of the three distinct groups, although Andelman has not been able to prove the result for all cases. If one of the three groups is electrically charged, the same result is likely, but strong interactions between oppositely charged groups bias the energetics towards chiral segregation in the crystal form.

The calculation of the phase diagrams is necessarily more difficult. To make the model correspond with some degree of realism to the case of molecules pinned to a two-dimensional layer but occupying what may be only a small part of it, while keeping the simplifying assumption that only three orientations of the tripod are allowed, Andelman is driven to a kind of lattice-gas model in which the vertices in a two-dimensional lattice are occupied by one of the enantiomers or by a vacancy.

Formally, this treatment is exactly like that of the simplest models of the adsorption of gases on the surfaces of solids. Despite continuing argument about the degree to which a two-dimensional structure, however well-ordered, can accurately represent a three-dimensional solid, two-dimensional lattice-gas models have at least the virtue of being exactly soluble.

The result is that Andelman is able to construct two generic phase diagrams, one each for the cases in which chiral segregation is preferred or not. The general shape of the diagrams resembles that of those for the sublimation of enantiomeric solids, whether racemic crystals or not, but there is a tendency for racemic crystals not to appear, but to be replaced by pure enantiomer, when the composition is too far from equality. Differences between this behaviour and that of real solids are not too disconcerting: two-dimensional structures are not solids, after all.

Part of Andelman's interest is to argue for the design of tripod molecules that could be studied in Langmuir-Blodgett films; it would certainly be interesting to know more about the structure of domains of chiral order. But there is probably still a lot of useful work to do in the elaboration of what seems to be a most useful model.

John Maddox

Steady steps lead to the gene

P. N. Goodfellow

IT WOULD be difficult to overstate the importance of the cloning of the cystic fibrosis gene (*CF*) which was reported in a series of papers last week in *Science*¹⁻³, and which is the successful culmination of a collaboration between the laboratories of L.-C. Tsui at the Hospital for Sick Children, Toronto, F.S. Collins at the University of Michigan and J.R. Riordan, at the University of Toronto.

Cystic fibrosis is a genetic disease that drastically reduces the quality of life and curtails life expectancy. In caucasian populations, 1 individual in 20 is a carrier of the defective gene. Consequently, in Britain, the birth incidence for this recessive disease is about 1 in 2,000. Afflicted individuals produce large amounts of thick mucus in their lung epithelia. This blocks the airways and promotes bacterial infection. Over the years, chronic lung infection results in gradual lung destruction and death is usually from this cause. Abnormal mucus production by epithelia elsewhere in the body may also cause complications, including the blockage of exocrine secretion. Treatments are designed to ameliorate symptoms and to prevent lung infection; cure is not possible⁴.

At the cellular level, previous research has identified an abnormality in ion transport in epithelial cells of cystic fibrosis patients⁵. The relationship between the ion transport defect and the underlying cause of cystic fibrosis is not known, nor is it clear what the relationship is between the ion-transport abnormalities and the production of thick mucus.

Two general strategies have been applied to isolating the *CF* gene. Theoretically, it should be possible to use physiological, pharmacological and biochemical methods to identify the defective protein in cells of cystic fibrosis patients. While making steady progress, this approach has been hindered by a general paucity of information about the structure and regulation of ion channels and associated molecules. The second strategy was to prove more fruitful; several groups, including the Tsui-Collins collaboration, have applied the package of molecular genetic techniques commonly referred to as reverse genetics. The principles are simple and have been propounded in these columns previously⁶. In brief, the target gene is first located on a chromosome by linkage analysis in family studies. Further genetic markers are sought that closely flank the target gene on both sides. Sequentially cloning all the sequences between flanking markers, a process known as chromosomal walking, must result in the cloning of the required gene. This leaves the final problem, which is to recognize the target

gene in the cloned sequences. The cloning of the *CF* gene closely followed this pathway.

In 1985, the first linkage between the *CF* gene and a genetic marker was found. Unfortunately, this linkage was to a gene (*PON*, encoding the polymorphic enzyme paraoxonase) that itself had not been located in the genome⁷. Soon afterwards, Tsui *et al.* localized *CF* to chromosome 7 by establishing linkage between *CF* and the random DNA segment *D7S15* derived from that chromosome⁸. The position on chromosome 7 was further refined by the discovery of the closely linked markers *MET* and *D7S8*, described by the groups of R. White and R. Williamson, respectively^{9,10}. Fortunately, *MET* and *D7S8* flanked the *CF* gene, thus defining a target region of approximately $1-2 \times 10^6$ base pairs in which it must lie. A chromosomal walk of this size is a massive undertaking. Therefore an attempt was made further to reduce the target size by screening for additional markers that mapped between *MET* and *D7S8*. Two were found — *D7S122* and *D7S340*. A third marker, *IRP*, isolated by Estivill *et al.* was originally proposed to be a candidate for the *CF* gene¹¹. But although it mapped very close to the *CF* gene, it became clear that it is distinct.

Armed with these probes, Tsui, Collins and colleagues initiated a series of bi-directional chromosome walks by isolating clones containing overlapping sequences derived from the target region. The efficiency of the walk was boosted by using chromosome jumping, a technique pioneered by Collins. Jumping serves two purposes: it allows the rapid production of additional starting points for chromosomal walks and prevents the walks from being stalled by unclonable sequences.

In all, about $2-3 \times 10^5$ base pairs of genomic sequence were cloned before the (5') end of the candidate *CF* gene was found. Several walks of this magnitude have been reported previously; nevertheless, few groups yet have the capacity for performing such a walk, and it is not easy — as testified by the description by Rommens *et al.*¹ of several unstable and unclonable sequences. As sequences were accumulated in the target region the end game was reached: it remained to identify the *CF* gene and to prove its causality for cystic fibrosis.

As Rommens *et al.* comment, there is no single or guaranteed route to gene identification. Experience gained from isolating other human genes, however, suggests that cross-species hybridization is a powerful approach. This provided the first clue for the existence of the candidate

CF gene; fortunately, although the cross-hybridization to rodent DNA was weak, bovine DNA proved better. Eventually, after screening seven different libraries with the bovine cross-reacting genomic clone, a single cDNA clone was isolated. Even at this point the problems were not over; Riordan *et al.*² had to isolate over 20 cDNA clones to assemble a complete open reading frame and many of these clones contained extraneous sequences or unspliced introns. Finally, it was necessary to use anchored polymerase chain reaction techniques to clone the untranslated sequences at either end of the coding sequence. These clones provided the tools to check the tissue distribution of the mRNA and to obtain the deduced amino-acid sequence of the candidate gene.

Was it the genuine *CF* gene? The tissue distribution of its mRNA was consistent with that predicted by the pathogenesis of cystic fibrosis. Also consistent was that the deduced amino-acid sequence indicated that the gene product is related to membrane transport proteins — for more on which, see opposite. This, and the need to avoid confusion with previous candidate *CF* genes, leads Riordan *et al.* to suggest that the protein is called cystic fibrosis transmembrane regulator (CFTR); it is to be hoped that the name has predicted the true function.

The best evidence that the *CF* gene had been isolated was obtained from correlating sequence data with genetic analysis^{2,3}. The major difficulty here was that there is strong linkage disequilibrium across the region where the *CF* gene is located. This means that it is difficult to distinguish between *CF* mutations and polymorphic variation showing very strong allelic association with the *CF* mutation. Nevertheless, about 70 per cent of the mutations in cystic fibrosis patients were associated with a deletion of a specific phenylalanine residue (ΔF_{508}). Moreover, this deletion was not found on chromosomes that did not carry a defective *CF* gene, even if they carried the same allele distribution as those that did. The allelic associations of the 30 per cent of mutations not due to ΔF_{508} suggests they are a heterogeneous group³. Final proof that the gene for CFTR corresponds to the *CF* gene will come if it can be shown in cell physiology experiments that the gene restores normal function to cystic fibrosis cells; the same system could be used to test the ΔF_{508} deletion.

The implications of this research are profound; there will be large spin offs in basic biology, especially in cell physiology, but the largest impact will be medical. A rough estimate would suggest that worldwide about 5,000 people die each year of cystic fibrosis. If the ΔF_{508} mutation alone is the major cause of the disease, detection of the 1 in 20 members of the population that carry the mutation could contribute

to the prevention of many of these deaths in the future. At present, prenatal diagnosis contributes to the avoidance of new cases of cystic fibrosis in families with the disease; the detection of carriers would make it possible to avoid the conception or birth of many currently unavoidable cases. I disagree with the sentiments of Kerem *et al.*³, who suggest waiting until all the mutations have been identified — carrier testing is worthwhile now.

Finally, there is the hope that knowledge of the structure and function of CFTR will lead to the design of rational approaches for treating cystic fibrosis. That hope, however, should be tempered with the lesson of sickle cell anaemia: molecular advances do not always lead to molecular cures¹². □

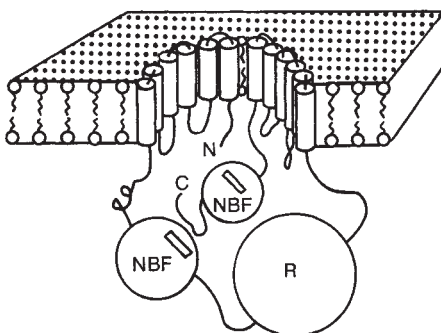
P.N. Goodfellow is in the Laboratory of Human Molecular Genetics, Imperial Cancer Research Fund, London WC2A 3PX, UK.

Protein joins transport family

Chris Higgins

OBTAINING a gene sequence is, in itself, not always especially informative. *CF* is clearly an exception¹, its sequence giving a strong indication of the biochemical nature of the cystic fibrosis defect. The CFTR protein that *CF* encodes shares considerable organizational and sequence similarities with a family of ATP-dependent transport systems that have been extensively studied in bacterial cells and which have more recently been shown to have eukaryotic counterparts (see ref. 13). The best characterized members of this family are the bacterial transporters referred to as the periplasmic or binding-protein-dependent transport systems. Over 20 such systems have been identified, each specific for a different substrate (such as an amino acid, sugar, peptide or inorganic ion)¹⁴. Transport requires the function of a 'periplasmic binding protein' (hence the family name) that delivers substrates to a complex of membrane proteins that mediate substrate translocation across the membrane.

The 'basic' membrane complex consists of four protein subunits (although there are variations on this theme), two of which are highly hydrophobic and span the membrane five or six times, and two of which are hydrophilic and associated with the cytoplasmic face of the membrane. The hydrophilic subunits share considerable amino-acid sequence similarity over much of their length. The conserved sequences include a potential ATP-binding site, and several of the proteins, indeed, bind ATP^{15,16}. Although none of the purified proteins has yet been shown to hydrolyse ATP, there seems little doubt that they do and stoichiometric ATP



Schematic model of CFTR, modified from ref. 2. The two membrane-spanning subunits are joined by a nucleotide(ATP)-binding fold (NBF) and the large R subunit. The other NBF is at the carboxy(C)-terminus.

hydrolysis concomitant with transport has been demonstrated¹⁷.

Several eukaryotic counterparts to these bacterial transporters have been identified. All differ from the bacterial transporters in that the individual protein subunits are combined into a single multidomain polypeptide, but the overall organization and amino-acid sequences are strongly conserved. The first eukaryotic example to be found was the MDR protein, or P-glycoprotein, which is responsible for multidrug resistance in tumours. The protein functions as a transporter, pumping drugs out of cells. The CQR protein responsible for chloroquine resistance in the malarial parasite turns out to be closely related to MDR¹⁸, as is the product of the *STE6* locus of yeast, which mediates export of α -factor mating pheromone from the cell^{19,20}.

It is to this widespread group of eukaryotic transporters that CFTR, the product of the *CF* gene, belongs. In its overall structure, CFTR is very similar to the other transport systems, having two hydrophobic domains, which comprise potential membrane-spanning helices, and two ATP-binding domains (see figure). Furthermore, CFTR includes an additional, large cytoplasmic domain (R in the figure) which, based on consensus phosphorylation sites for protein kinase C and cAMP-dependent protein kinase, probably serves a regulatory role. Thus there seems little doubt that CFTR is a transport system, but for what?

It has become clear that cystic fibrosis patients show altered chloride transport across epithelia and that this is the root cause of the most severe clinical symptom — decreased fluid secretion and accumulation of dehydrated mucus in the bronchial tubes. Furthermore, this chloride channel is subject to complex regulation, only now being unravelled. Channel activity is increased by cAMP-dependent protein kinase in normal subjects but not in patients; activation by protein kinase C also seems to be defective in patients²¹⁻²⁴. Biochemically it is unclear whether the primary defect is in the chloride channel

or in another component of the complex regulatory network. The similarity of CFTR to other membrane transporters strongly suggests that it is part of the chloride transporter, and hence it is this pump that is altered in cystic fibrosis patients. But this has yet to be demonstrated. It may be significant that the predicted molecular mass of CFTR is rather different from that of functional chloride channels recently isolated from epithelia²⁵.

Over 60 per cent of cystic fibrosis patients seem to have the same mutation, a deletion of three nucleotides, removing a single phenylalanine residue from CFTR². This amino acid is in one of the highly conserved ATP-binding domains of the protein but is not one of the supposedly crucial amino acids in the 'core' site¹⁶. Does the mutation completely destroy the function of the CFTR or is it a specific missense mutation that alters the protein's activity? There is support for the latter view as patients' regulation of chloride flux is altered rather than absent.

If the cystic fibrosis phenotype involves protein misfunction, not loss of function, there is a better chance of devising a corrective treatment for the disease; it would be hard to replace an absent transport protein, but it may be possible to modulate the activity of one that is perturbed. In what way might such a missense mutation affect the activity of the CFTR protein? One possibility is a subtle effect on the coupling of ATP hydrolysis to the transport process. An understanding of this will probably require determination of the three-dimensional structure of these sites, not impossible given recent progress on bacterial transport proteins. Their ATP-binding domains are so similar to CFTR that there is no doubt that they provide an important structural model. □

Chris Higgins is in the ICRF Laboratories, Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DU, UK.

- Rommens, J.M. *et al.* *Science* **245**, 1059–1065 (1989).
- Riordan, J.R. *et al.* *Science* **245**, 1066–1073 (1989).
- Kerem, B.-S. *et al.* *Science* **245**, 1073–1080 (1989).
- Dodge, J.A. in *Cystic Fibrosis* (ed. Goodfellow, P.N.) 12–23 (Oxford University Press, 1989).
- Knowles, M.R. *et al.* *Science* **221**, 1067–1070 (1983).
- Goodfellow, P.N. *Nature* **322**, 12 (1986); **326**, 824 (1987).
- Eiberg, H. *et al.* *Cytogenet. Cell Genet.* **40**, 623 (1985).
- Tsui, L.-C. *et al.* *Science* **230**, 1054–1057 (1985).
- Knowlton, R.G. *et al.* *Nature* **318**, 380–382 (1985).
- Wainwright, B.J. *et al.* *Nature* **318**, 384–385 (1985).
- Estivill, X. *et al.* *Nature* **326**, 840–845 (1987).
- Luzzatto, L. & Goodfellow, P.N. *Nature* **337**, 17 (1989).
- Higgins, C.F. *Nature* **340**, 342 (1989).
- Ames, G.F.-L. *A. Rev. Biochem.* **55**, 397–425 (1986).
- Higgins, C.F. *et al.* *EMBO J.* **4**, 1033–1040 (1985).
- Higgins, C.F. *et al.* *Bioessays* **8**, 111–116 (1988).
- Mimmack, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Foot, S.J. *et al.* *Cell* **57**, 921–930 (1989).
- McGrath, J.P. & Varshavsky, A. *Nature* **340**, 400 (1989).
- Kuchier, K. *et al.* *Cell* (in the press).
- Widdicombe, J.H. *et al.* *Proc. nat. Acad. Sci. U.S.A.* **82**, 6167 (1985).
- Schoumacher, R.A. *et al.* *Nature* **330**, 752 (1987).
- Hwang, T.-C. *et al.* *Science* **244**, 1357 (1989).
- Li, M. *et al.* *Science* **244**, 1353 (1989).
- Landry, D.W. *et al.* *Science* **244**, 1469–1472 (1989).

Wetter clouds dampen global greenhouse warming

Tony Slingo

VARIATIONS in cloudiness can significantly affect the Earth's radiation budget — the absorbed solar energy and the thermal energy radiated back to space¹. This conclusion is supported by various modelling studies of the 'greenhouse' effects of increased CO₂ concentrations²⁻⁴. In these studies, the CO₂ forcing induces changes in the cloud coverage and liquid-water content, which alter the radiation budget further and thus modify the original perturbation. On page 132 of this issue⁵, Mitchell *et al.* add a new twist to the story, showing how changes in the relative amounts of ice-crystal and water-drop clouds substantially reduce the sensitivity of their model to changes in concentration of CO₂. Indeed, it is becoming apparent that uncertainties in the treatment of clouds severely undermine model prediction of climate.

Mitchell *et al.* used various versions of the UK Meteorological Office atmospheric general circulation model⁶ coupled to a 50-m ocean mixed layer⁷. Replacing the standard cloud prediction scheme based on relative humidity by a more sophisticated prescription including liquid water and ice virtually halves the average equilibrium surface warming due to doubled CO₂ — from 5.2 K to 2.7 K. They argue that this spectacular sensitivity to details of the cloud scheme is caused by the replacement of ice clouds by water clouds near the freezing level. In the new cloud scheme, the fall-speed assumed for water drops is much smaller than that for ice crystals, so that liquid-water clouds can dissipate more slowly than ice clouds. According to Mitchell *et al.*, this leads to the predicted relative increase in water clouds near the freezing level, which consequently reduces the amount of solar energy absorbed in the system (more is reflected back into space) and hence gives a smaller warming. When, in addition, cloud radiative properties are allowed to depend on the liquid-water content of clouds, the surface warming drops still further, to only 1.9 K.

The strong dependence of the magnitude of cloud feedback in this work on the details of the model parallels results presented recently⁸ by a large *ad hoc* group (including Mitchell and myself) led by Robert Cess of the State University of New York. We compared 14 general circulation models and found a three-fold variation in global climate sensitivity, caused largely by differences in the cloud feedbacks. The models disagreed not only as to the magnitude but also the sign of the net cloud feedback. For this reason, it is

clear that uncertainty in how to formulate clouds represents a formidable obstacle to reliable climate prediction.

This underlines the continuing need for physical parameterizations to be compared with results from field studies. This is particularly important now that models are becoming increasingly complex. Mitchell *et al.* argue that the assumptions in their model are 'at least qualitatively realistic', but if climate change simulations are to be believable, then all parts of a model must be validated quantitatively, and important processes that are ignored must be identified. Of particular relevance to the cloud problem is the FIRE programme (first regional experiment of the international satellite cloud climatology project), both in its stratocumulus⁹ and cirrus¹⁰ field phases. Several papers are in preparation that highlight the complexity of the meso-scale dynamical structure of cirrus, so that the fall-speed of ice crystals is only part of the story. It is also interesting that two cases of highly supercooled liquid-water altostratus clouds at temperatures around -30 °C were observed (A.J. Heymsfield *et al.*, personal communication), which suggests that the transition from water to ice clouds is more subtle than Mitchell *et al.* assumed for their model.

The substantial sensitivity of climate models to the details of their formulation,

ECOLOGY

Just proportions in food webs

Joel E. Cohen

IN FOOD webs of natural communities with varying numbers of species, the number of green plants and other species that consume no prey in the web (basal species) is nearly proportional, on average, to the number of species that have both predators and prey in the web (intermediate species), according to studies reported over the past dozen years¹⁻⁴. Empirical rules^{5,6} of that type describe structural or anatomical properties of food webs. An important question in community ecology is whether parallel regularities exist in the way in which food webs function^{7,8}. On page 142 of this issue⁹, McNaughton *et al.* report remarkable regularities in measures of food web function. Their findings indicate the existence of a physiology of food webs that generalizes across natural communities. Their report opens up the distant, but highly attractive, prospect of linking the previously known structural regularities

and the fact that 14 models give 14 different answers for the cloud feedback, show that we are far from the goal of accurate predictions of future climate change. Sensitivity studies such as that reported by Mitchell *et al.* and comparisons between models as done by Cess *et al.* are now the bread and butter of climate modelling. But at the present rate of progress, it will be several years before reliable predictions of global and regional climate change are available from the models. That underlines the need for comprehensive monitoring that is capable of unravelling the climate change signal from the noise and of providing enough data on the state of the atmosphere, the Earth's radiation budget and clouds for the feedbacks to be determined by observation. Such an analysis may well yield the fifteenth answer but at least it will have come from the real world and will permit a much needed test of the models. □

Tony Slingo is in the Climate and Global Dynamics Division of the National Center for Atmospheric Research, PO Box 3000, Boulder, Colorado 80307, USA.

1. Ramanathan, V. *et al.* *Science* **243**, 57–63 (1989).
2. Schlesinger, M.E. & Mitchell, J.F.B. *Rev. Geophys.* **25**, 760–798 (1987).
3. Schlesinger, M.E. *Nature* **335**, 303–304 (1988).
4. Roeckner, E. *Nature* **335**, 304 (1988).
5. Mitchell, J.F.B., Senior, C.A. & Ingram, W.J. *Nature* **341**, 132–134 (1989).
6. Slingo, A., Wilderspin, R.C. & Smith, R.N.B. *J. geophys. Res.* **94**, 2281–2301 (1989).
7. Wilson, C.A. & Mitchell, J.F.B. *J. geophys. Res.* **92**, 13315–13343 (1987).
8. Cess, R.D. *et al.* *Science* **245**, 513–516 (1989).
9. Albrecht, B.A., Randall, D.A. & Nicholls, S. *Bull. Am. met. Soc.* **69**, 618–626 (1988).
10. Starr, D.O. *Bull. Am. met. Soc.* **68**, 119–124 (1987).

with new functional patterns.

McNaughton *et al.* show that the biomass, consumption and net secondary productivity (NSP) of herbivores are all related to the net above-ground primary productivity (NAP) of a natural habitat by simple power laws. Particularly striking is that the exponent of the power law that relates NSP to NAP is not significantly different from 1: to a first approximation, NSP and NAP are directly proportional. In effect, the energy converted into herbivores per square metre of surface area per year is, on average, about one-thousandth of the energy that is converted into plants per square metre of surface area per year. The ecological energy pyramid has a much larger base than second storey.

This proportionality would be a triviality if secondary productivity were passively controlled and limited by primary productivity. But this trickle-down view of

community structure ignores evidence that herbivores not only depend on, but strongly influence, the plants they eat. In *The Origin of Species*, Darwin reported that, on a small cleared plot, he "marked all the seedlings of our native weeds as they came up, and out of 357 no less than 295 were destroyed, chiefly by slugs and insects." More recent studies show clearly that the feedback effects of herbivores on plant populations are both negative¹⁰ and positive¹¹. In the light of these studies, it is far from an expected finding that the complex interplays between plants and herbivores should yield a simple proportionality between NAP and NSP. Moreover, the constant of proportionality, one-thousandth, is far less than might have been guessed from widely quoted ecological efficiencies of the order of 10–20 per cent. Nor would the trickle-down theory explain why consumption by herbivores rises as the square of net foliage production, as McNaughton *et al.* find.

If studies of food-web structure demonstrate a rough proportionality between the numbers of basal and intermediate species (trophic species, really, since biological species with identical predator species and identical prey species are counted as a single unit), and if future studies of food-web function confirm the finding of McNaughton *et al.* that NSP and NAP are proportional, the next problem, and opportunity, is to relate the two proportionalities. In the 35 studies where NSP and NAP were both measured directly, are the numbers of basal trophic species proportional to the numbers of herbivore trophic species? What is the relation between primary or secondary productivity and the diversity of biological and trophic species in these studies? Stay tuned!

But as Mark Twain advised a young reporter: first get your facts straight, then you can distort 'em. McNaughton *et al.* acknowledge that their findings would be considerably strengthened by more data from tropical forests and other structurally complex systems. At present, their results rely heavily on structurally simple vegetation such as grassland and tundra. Moreover, the wide scatter of the assembled data with respect to the log-log linear regression proposed to describe them brings to mind the statistical maxim that, on log-log coordinates, even an elephant looks like a straight line. The report of McNaughton *et al.* does not have all the answers, but it has some and, better still, it raises important new questions.

Both the structural and the functional proportionalities found in food webs are purely descriptive. Though a model exists that explains much of food-web structure^{5,12}, it too is phenomenological; and there is no hint of a mechanistic explanation for the observed patterns in the report of McNaughton *et al.* But ecology is a

very young science. A backward glance to the early stages of other, now maturer, sciences shows the tremendous power of empirical proportionalities. Joseph Louis Proust proved that elements combine in definite proportions by weight when they form chemical compounds (the law of definite composition), laying a cornerstone of the empirical support for the atomic theory¹³. Gregor Mendel found fixed proportions of different phenotypes in offspring in his experiments on plant breeding; his explanation in terms of genes was phenomenological. Erwin

Chargaff's discovery that the concentration of adenine equals that of thymine and that the concentration of guanine equals that of cytosine in DNA from various sources was one of the chief clues to Watson and Crick's solution of DNA structure. Time will reveal the scientific consequences of the new proportionalities in food-web structure and function. The possibilities are exciting. □

Joel E. Cohen is Professor of Populations at The Rockefeller University, New York, New York, 10021–6399, USA.

1. Cohen, J. *Nature* **270**, 165–166 (1977).
2. Briand, F. & Cohen, J. *Nature* **307**, 264–267 (1984).
3. Jeffries, M.J. & Lawton, J.H. *Freshwater Biology* **15**, 105–112 (1985).
4. Sugihara, G., Schoenly, K. & Trombla, A. *Science* **245**, 48–52 (1989).
5. Cohen, J.E., Briand, F. & Newman, C.M. *Community Food Webs: Data and Theory* Biomathematics Vol. 20 (Springer, Heidelberg, 1989).
6. Lawton, J.H. & Warren, P.H. *Trends Ecol. Evol.* **3**, 242–245 (1988).
7. Pimm, S. *Food Webs* (Chapman and Hall, London, 1982).
8. Cole, J.J., Findlay, S.E.G. & Lovett, G.M. (eds) *Compara-*

- tive Analyses of Ecosystems: Patterns, Mechanisms and Theories* (Springer, New York, 1990).
9. McNaughton, S.J., Oesterheld, M., Frank, D.A. & Williams, K.J. *Nature* **341**, 142–144 (1989).
10. Dirzo, R. in *Perspectives on Plant Population Ecology* (eds Dirzo, R. & Sarukhán, J.) 141–165 (Sinauer, Sunderland, Massachusetts, 1984).
11. Bianchi, T.S., Jones, C.G. & Shachak, M. *Trends Ecol. Evol.* **4**, 234–238 (1989).
12. Cohen, J.E. & Newman, C.M. *Proc. R. Soc. B* **224**, 421–448 (1985).
13. Jaffe, B. *Crucibles: The Story of Chemistry* (Fawcett, New York, 1957).

ASTRONOMY

Dark secrets of a giant galaxy

M. G. Edmunds

THE discovery of a single exceptional object can radically affect accepted ideas. If a single man more than 60 feet in height were suddenly to be found in, say, Outer Mongolia there would no doubt be an urgent re-examination of current opinion in human genetics. Although the sizes of galaxies span a much greater range than those of people, recently published observations by C. Impey & G. Bothun (*Astrophys. J.* **341**, 89–104; 1989) of an exceptionally large galaxy, suggest a need for careful reappraisal of current opinions on the formation and evolution of spiral galaxies.

The exceptional object is Malin 1, discovered by chance two years ago and named after David Malin, who pioneered the photographic amplification technique crucial in its detection (G. D. Bothun, C. Impey, D. F. Malin & J. R. Mould *Astr. J.* **94**, 23–29; 1987). Amplification was necessary because the surface brightness — the brightness per unit area — of the galaxy on the sky is very small: when fighting against the light of the night sky (which is not nearly as dark as it appears to the naked eye), an extended object of low surface brightness is hard to find. Indeed, recognition of Malin 1 was possible only because it has a fairly bright central region (bulge component and nucleus) showing a few weak emission lines. These could be used to measure a redshift, so giving a distance and allowing the precise tuning of radio telescopes to investigate the hydrogen content.

What emerged was a galaxy with over three times as much neutral hydrogen

as any galaxy previously reported, and with an extraordinary physical size. The dim disk has an exponential scale length (the length over which the brightness decreases by a factor of *e*) of at least 55 kiloparsecs (kpc). The total extent in neutral hydrogen is greater than 240 kpc. The exponential scale length should be compared with that of our own Galaxy (5 kpc) and Andromeda (6 kpc), themselves fairly large relative to typical spirals, whose scale lengths are around 2 kpc.

Despite its size, it was the faintness of the disk that had made Malin 1 inconspicuous. Two questions obviously spring to mind. First, what effect does meagre surface brightness — and, by direct implication, low physical density of material — have on the evolution of a galaxy, and second, are there many more such galaxies out there?

The new spectroscopic observations, by Impey and Bothun, of Malin 1 show that the central bulge component is not particularly unusual. It displays the features expected of an old and evolved stellar population, much as one might find in an average spiral, with some evidence of continuing star formation. It is the disk that is unusual. This prompts the authors to raise the question whether the bulge and disk may have been essentially independent of each other since they were formed. The density of the disk is so low that it is near to a theoretical limit for the growth of instabilities which could lead to star formation.

So the disk may have been just sitting there, with little happening, since it formed.

Malin 1 is situated in a region of comparatively empty space, and there may have been no encounters with other galaxies to disturb its fragile, mainly gaseous, disk. Alternatively, the disk could be an addition, perhaps recently accreted by the central galaxy from the intergalactic medium and not yet sufficiently dense to have formed many stars.

The first idea is attractive because it agrees with several speculations about the origin of some of the absorption lines due to the Lyman- α series of hydrogen that are a feature of the measured spectra of some distant quasars (see, for example, A. M. Wolfe, D. A. Turnshek, H. E. Smith & R. D. Cohen *Astrophys. J. Suppl.* **61**, 249–304; 1986). The strength of the absorption lines indicates high column-densities of neutral hydrogen in the absorbing objects lying between us and the quasars. Indeed, the measurements suggest that the properties of those objects match those of disks of low surface brightness, such as Malin 1.

What the quasar measurements show is that absorbing objects with strong Lyman- α lines are reasonably common at redshifts of about 2. If they are indeed disks on the pattern of Malin 1, which is relatively nearby, the implication is that it must be possible for disks of this kind to remain stable for periods of time corresponding to the interval between redshift 2 and (almost) the present. That suggests that large tenuous gas disks may still be fairly numerous — indeed, Impey and Bothun suggest that it would be consistent with their observations that the mass of disks such as Malin 1, still mostly unobserved, might amount to as much as half the mass of the known galaxies.

But that constrains galactic evolution to a fairly tight schedule. If a substantial proportion of the disks have remained unchanged since redshift 2, it is not easy to understand how there could have been time enough for them to have been formed in large numbers by that time. Although the interval in which galaxies were formed remains uncertain, it was probably between redshifts 10 and 3. According to Wolfe *et al.*, that implies the disks could only have been between 1,000 and 4,000 million years old by the time they are observed in the quasar absorption spectra at redshift 2. This is the kind of timescale now used in models of the formation of galaxies by the collapse of gas clouds.

It will be important if other extended disks of low-surface brightness like Malin 1 can be found, to show that galaxies *can* have well-developed disks at an early stage. (Such observations would also show that Malin 1 is not just a rare, late developer.) The importance of such objects for theories of the evolution of spiral galaxies would be in increasing the ranges of both the physical size and the evolutionary rates to be explained. They may show that

spiral galaxies form quickly as individual gaseous-disk systems, but that the subsequent rate of evolution (as judged by how fast the gas turns into stars) varies much more than previously realized, depending critically on parameters such as the density of material.

Although there are other large galaxies with low surface brightness, Malin 1 is by far the largest known. All have active nuclei, which is probably why they have been identified. It is not easy to predict how many more await discovery, especially if they are inconspicuous without bright centres. Even nearby systems could have escaped notice. Impey and Bothun point out that, because of the characteristics of radio telescopes, Malin 1 might not have been recognized had it been as close as the

Virgo Cluster, behind which it was found.

Nearby examples should be eagerly sought. It would be invaluable to learn more about their properties, particularly the age and nature of the stellar population in the disks and their total masses and mass distributions from rotation curves. But even for nearby examples, the observational difficulties will be very great because the surface brightness (which is what essentially determines the quality of observations for an extended object) does not change with distance, and it will also be (by definition) low for nearby examples of these galaxies. □

M. G. Edmunds is in the Department of Physics, University of Wales, PO Box 913, Cardiff CF1 3TH, UK.

LIPOPROTEIN-REMNANT RECEPTORS

Second receptor verified?

Anne K. Soutar

CHYLOMICRON remnants, the lipoprotein by-products of dietary fat absorption and transport, are rapidly and efficiently cleared from the circulation by a mechanism located exclusively in the liver¹, but how? Although the receptor for low-density lipoprotein (LDL) recognizes chylomicron remnants and could have that function², there must be a second genetically distinct mechanism to account for the apparently normal clearance of chylomicron remnants in patients with familial hypercholesterolaemia³ (FH) and in WHHL rabbits⁴, both of which lack functional LDL receptors. Now, two papers, one by Beisiegel *et al.* on page 162 of this issue⁵ and another from Brown and Goldstein's laboratory in Dallas⁶ go some way to strengthen the belief that the latest candidate for this role — LRP, otherwise LDL-receptor-related protein — may indeed be the second receptor. But the issue is not yet settled, and LRP may have other interesting functions as well, or even instead.

The complementary DNA of LRP was found⁷ last year by screening a mouse lymphocyte cDNA library for sequences corresponding to the highly conserved cysteine-rich motif of the ligand-binding domain of the LDL receptor, also present in the terminal components of complement. The predicted amino-acid sequence of one selected clone turned out to be quite remarkably similar to that of the LDL receptor⁸ (see figure), raising the possibility that it might be the previously elusive chylomicron-remnant receptor. The homology also suggested absorbing questions about its evolutionary relationships with other proteins, notably the LDL receptor⁹ itself and the precursor of epidermal growth factor. But it is too soon to conclude that LRP is indeed the

chylomicron-remnant receptor molecule.

Uptake experiments *in vivo* and in perfused liver show that there are two functional features that should distinguish the chylomicron-remnant receptor from the LDL receptor: it should recognize lipoproteins that contain apoE, but not those such as LDL that contain only apoB; and its level of expression should not be influenced by intracellular cholesterol accumulation. Otherwise, the receptor should resemble the LDL receptor in its location on the surfaces of liver cells and in its capacity to mediate endocytosis and lysosomal degradation of whole lipoprotein particles.

The new paper by Beisiegel *et al.* is the result of collaboration between the Hamburg group and Herz and Stanley at the EMBL at Heidelberg, who originally cloned LRP. It provides fairly convincing evidence that LRP is a cell-surface protein that binds apoE. This they have done by the use of phospholipid liposomes complexed with apoE that has been chemically modified with a crosslinking agent capable of being photoactivated. They have shown that if these liposomes are incubated either with HepG2 cells or with isolated human liver membranes, the crosslinker labels a single-membrane protein of large molecular weight that is immunologically indistinguishable from LRP.

But these observations alone do not, of course, establish identity. Because the liposomes used are by no means identical with authentic chylomicron remnants, whose use would have complicated the study because they contain apoB and apoC as well as apoE, it was important for the authors to demonstrate that there is competition for binding sites between the artificial liposomes and native lipoproteins containing apoE. However, they

have not further investigated the specificity of the receptor binding. In particular, they have not applied the essential test of showing that the receptor does not bind LDL, but it would also have been interesting to know whether lactoferrin, which specifically inhibits¹⁰ hepatic uptake of chylomicron remnants but not LDL, interferes with apoE binding to LRP. And not everybody will agree with the authors' claim that the equal affinity for LRP of apoE₂ (an isoform of apoE that does not bind to the LDL receptor) and the other isoforms of apoE is further evidence that LRP is the chylomicron-remnant receptor.

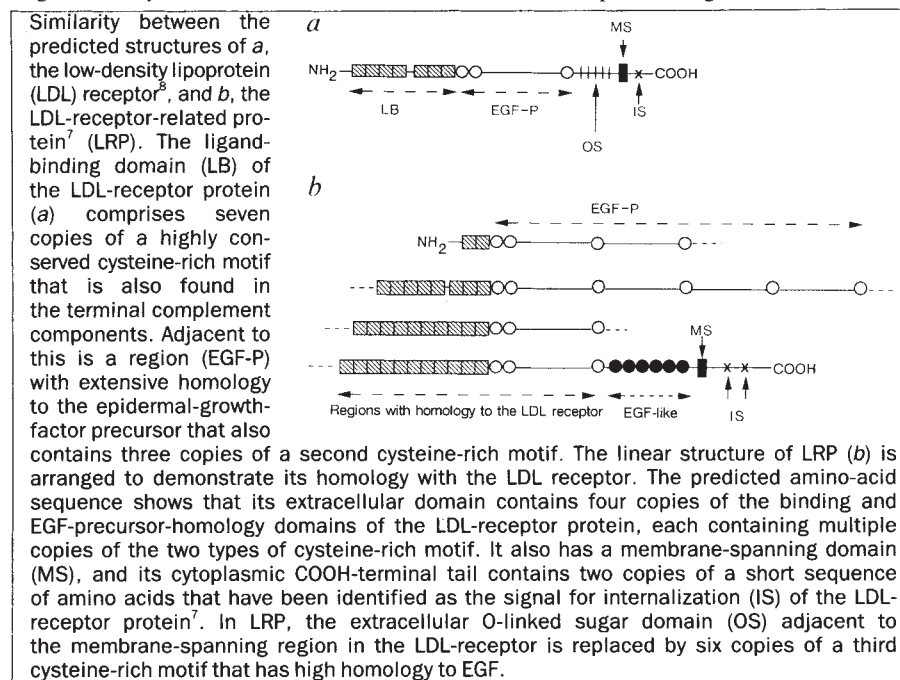
Evidence of a different nature, from Brown and Goldstein's group⁶ at Dallas (now joined by Herz from Heidelberg), shows that in skin fibroblasts from an FH patient unable to synthesize LDL receptor molecules, LRP mediates the uptake and lysosomal degradation of cholesteryl esters derived from lipoproteins enriched in apoE. The authors have not been able to demonstrate that LRP directly determines either the binding or degradation of radiolabelled lipoproteins, but instead have followed changes in the incorporation of ¹⁴C-labelled oleate into cholesteryl esters. The assumption is that, if the rate of incorporation is greater when cells are incubated with lipoprotein, the lipoproteins themselves must have been subjected to endocytosis mediated by receptors and to degradation by lysosomes, with the release of free cholesterol.

This argument is not watertight, but the conclusion that lysosomes are involved is supported by the observed inhibition of oleate incorporation by chloroquine. A further difficulty is that it has not been possible to demonstrate these changes in oleate incorporation with native lipoproteins. Although the authors worked with rabbit β -VLDL (very-low-density lipoprotein), which accumulates in the plasma of rabbits fed cholesterol, they were unable to demonstrate LRP-dependent stimulation of oleate incorporation unless they pre-incubated the β -VLDL with exogenous apoE before adding it to the cells. This is surprising, because β -VLDL is already apoE rich and binds with high affinity to the LDL receptor. But changes of oleate incorporation were suppressed by pre-incubating the cells with anti-LRP polyclonal IgG, indicating that intracellular events are influenced by some interaction between apoE and LRP even if the evidence is not compelling that LRP mediates endocytosis of the whole lipoprotein.

Further circumstantial evidence that LRP is indeed the chylomicron-remnant receptor comes from the structure of the gene and its genetic environment. The promoter region of the LRP gene does not contain a sterol regulatory element of the kind associated with the LDL receptor (as

well as with the genes of HMG CoA reductase and HMG CoA synthase), which agrees with the observation that the expression of LRP is unaffected by sterols¹¹. Yet the most obvious difficulty in equating LRP and the chylomicron-remnant receptor remains its tissue distribution. The protein is expressed in several tissues and is, for example, as abundant in brain and lung as in liver⁴. Yet there is no evidence that chylomicron remnants are degraded anywhere but in the liver. It

function? Stanley and his colleagues offer¹¹ as a clue the observation that an early degradation product of LRP is a protein cluster of relative molecular mass of roughly 90,000, which is consistent with the notion that the four LDL-receptor-like domains have been specifically cleaved away, leaving only the membrane-bound fragments of the molecule which, it turns out, are remarkably homologous (see figure) with the precursor of epidermal growth factor. This



would be surprising if LRP had no function in other tissues.

Meanwhile Stanley, who has moved to the Heart Research Institute in Sydney, has continued to explore other biological functions of LRP as well as its putative role as a remnant receptor. He reported at a recent meeting* that, for such a large molecule, LRP turns over rapidly in cultured cells and that it differs from the LDL receptor in not recycling to the surface after internalization. Interestingly, the short amino-acid sequence in the cytoplasmic tail of the LDL receptor required for clustering and endocytosis⁹ serves the same function in LRP. But there is a second copy of this sequence in the cytoplasmic tail of LRP and it has been shown (by substituting the cytoplasmic tail of LRP for the corresponding structure of haemagglutinin A in the influenza virus) that the presence of the second copy is both necessary and sufficient to ensure rapid degradation of the protein. Plainly we shall learn as much about the regulation of intracellular protein traffic from LRP as we have done from the LDL receptor.

But if LRP proves not to be the chylomicron-remnant receptor, what can be its

truncated LRP could thus serve to release growth factors, but Stanley (personal communication) also believes that, because LRP is found chiefly in cells synthesizing apoE¹², its function may be related to the secretion of this lipoprotein. If one also recalls that apoE synthesis is increased in regenerating nerve fibres¹², the prospects are that the functions of LRP will extend far beyond that of simply binding chylomicron remnants. □

Anne K. Soutar is in the MRC Lipoprotein Team, Hammersmith Hospital, London W12 0HS, UK.

1. Redgrave, T.G. *J. clin. Invest.* **49**, 465–471 (1970).
2. Wade, D.P., Knight, B.L. & Soutar, A.K. *Eur. J. Biochem.* **159**, 333–340 (1986).
3. Fredrickson, D.S., Goldstein, J.L. & Brown, M.S. in *Metabolic Basis of Inherited Disease* 4th edn (eds Stanbury, J.B., Wyngaarden, J.B. & Fredrickson, D.S.) 604–655 (McGraw-Hill, New York, 1978).
4. Kita, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 3623–3627 (1982).
5. Beisiegel, U., Weber, W., Ihrke, G., Herz, J. & Stanley, K.K. *Nature* **341**, 162–164 (1989).
6. Kowal, R.C., Herz, J., Goldstein, J.L., Esser, V. & Brown, M.S. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5810–5814 (1989).
7. Herz, J. *et al. EMBO J.* **7**, 4119–4127 (1988).
8. Sudhof, T.C., Goldstein, J.L., Brown, M.S. & Russell, D.W. *Science* **228**, 815–822 (1985).
9. Davis, C.G., van Driel, I.R., Russell, D.W., Brown, M.S. & Goldstein, J.L. *J. biol. Chem.* **262**, 4075–4082 (1987).
10. Huettlinger, M., Retsek, H., Eder, M. & Goldenberg, H. *Clin. Biochem.* **21**, 87–92 (1988).
11. Kütt, H., Herz, J. & Stanley, K.K. *Biochem. Biophys. Acta* (in the press).
12. Mahley, R.W. *Science* **240**, 622–630 (1988).

* 53rd annual meeting of the European Artherosclerosis Society, Vienna, 20–22 April 1989.

A primitive immune system

Charles A. Janeway

THE first great controversy in immunology pitted Metchnikov and the cellularists against Robert Koch and the humoralists¹. One consequence of the subsequent humoralist victory was the focus of immunology on the chemistry of antibodies and the abandonment of inquiry into the cells that mediate host defence. Interest in some of these cells was revived by Burnet's clonal selection theory of antibody formation² and by Gowans's discovery that lymphocytes mediate all specific immune responses³. These milestones in immunology directed attention toward the specificity of antigen recognition, and away from Metchnikov's original interest in primitive host cellular responses to infection.

As our understanding of specific antigen recognition has exploded, culminating in the discovery of seven families of rearranging genes encoding T- and B-lymphocyte receptors⁴, one can start to ask a more metchnikovian question: did immunological effector mechanisms evolve after the sophisticated recognition systems that now trigger them, or are the recognition systems recent additions to pre-existing, non-clonal systems of host defence⁵? Natural killer (NK) cells — lymphocytes (non-T, non-B) that can kill certain tumour cells^{6,7} — may constitute such a system. In demonstrating the presence of the invariant ζ -chain component of the T-cell receptor in NK cells, the paper of Anderson *et al.* on page 159 of this issue⁸, may help to answer the question.

Host defence against infection occurs in three phases in mammals: innate resistance that is not inducible; an early, inducible phase that is largely antigen-nonspecific; and a late, T-cell dependent phase that is highly antigen specific and generates immunological memory⁹. NK cells participate in the innate¹⁰ and early, interferon-inducible^{11,12} phases of the immune response to viral infection (see table below).

Viruses parasitize host cells, where they

replicate and produce new virions that spread the infection. The immune response blocks viral spread by making specific neutralizing antibodies and specific (class I MHC-restricted) cytolytic T cells that kill virus-infected cells. Although NK cells are identified by their ability to kill certain tumour cells without prior immunization^{6,7}, recent evidence has suggested a more important role of NK cells in host defence against viral infection¹³. This is supported by the discovery of an individual whose only demonstrable immunological defect was the absence of NK cells¹⁰. The early phases of herpes virus infections were markedly exaggerated in this patient and would have been fatal without anti-viral and supportive therapy, although she ultimately made antibody and cytolytic T-cell responses that cleared the infections.

How does an NK cell that lacks a clonally distributed receptor know which target cell to attack? The answer to this question remains shrouded in mystery; indeed, it may be the wrong question. Rather, a single NK cell may have multiple cognitive mechanisms. Thus, NK cells may function as a back-up defence mechanism for cytolytic T cells by destroying cells that lack class I MHC molecules¹⁴; they have Fc receptors that allow them to kill antibody-coated target cells¹⁵; they clearly kill certain virus-infected cells^{10,13}; they appear to regulate haemopoiesis (through recognition of structures mapping between H-2S and H-2D in the MHC of mice)¹⁶; and they can directly kill certain bacteria¹⁷.

That NK cells and cytolytic T cells share cytolytic effector mechanisms suggests that the NK cell is an evolutionary forerunner of the cytolytic T cell, just as the 'alternative' pathway of complement activation is likely to be a forerunner of the 'classical' pathway triggered by specific antibody (note the bias in the naming of the 'classical' pathway by humoralists). It seems likely that the host defence systems of primitive organisms consisted of effector cells like the NK cell

with multiple non-clonally distributed cognitive mechanisms. Primitive immune systems would be expected to have had few cells, so the effector cells must have been multi-specific. The acquisition by an NK cell of a clonally distributed receptor capable of fine discrimination between self and non-self might have been sufficient to

make it a cytolytic T cell. Clonally distributed receptors allow an immune system to recognize infection by any virus and the development of immunological memory through clonal selection.

But the T-cell receptor is not just an antigen-binding molecule. It is also the ligand-binding domain of a signal-transducing molecular complex that includes the CD3 proteins¹⁸ and a polypeptide known as the ζ -chain¹⁹, which is required for effective signal transduction by the T-cell receptor²⁰. Homologous chains are also associated with the signal-transducing Fc_γ receptor of the mast cell²¹. Thus ζ -like chains may be a common component of signal transduction mechanisms in cells of haemopoietic origin.

The identification by Anderson *et al.*⁸ of NK cell ζ -chains, associated with larger structures that may participate in ligand recognition, supports the idea that cytolytic T cells arose from NK-like cells by acquiring a clonally distributed T-cell receptor. Further support for this notion comes from finding NK-like cytotoxic activity in long-term cultured T-cell lines²². Such discoveries should help in understanding the evolution of host defence mechanisms. They could spur a reunification of the metchnikovian interest in innate immunity with the mainstream of specific recognition first championed by the humoralists. It is interesting to wonder whether progress has been speeded by the triumph of the humoralists that led to the solution of the antibody problem more than it has been impeded by the early abandonment of metchnikovian thinking. □

Charles A. Janeway is at the Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06510, USA.

The three phases of host defence against viral infection

Phase	Characteristics	Mechanism
Immediate (< 4 h)	Nonspecific, innate, no memory, no specific T cells	Natural killer cells
Early (4–96 h)	Nonspecific and specific, inducible, no memory, no specific T cells	Interferons α and β Interferon-activated natural killer cells
Late (> 96 h)	Specific, inducible, memory, specific T cells	Cytolytic T cells Interferon Specific antibody

1. Silverstein, A.M. *A History of Immunology* (Academic, San Diego, 1989).
2. Burnet, F.M. *The Clonal Selection Theory of Immunity* (Vanderbilt University Press, Nashville, 1959).
3. Gowans, J.L. & McGregor, D.D. *Prog. Allergy* **9**, 1 (1965).
4. Tonegawa, S. *In vitro cell. dev. Biol.* **24**, 253–265 (1988).
5. Janeway, C.A. Jr *Cold Spring Harb. Symp. quant. Biol.* **54**, 112–116 (1975).
6. Kiessling, R., Klein, E. & Wigzell, H. *Eur. J. Immun.* **5**, 112–116 (1975).
7. Herberman, R.B., Nunn, M.E. & Lavrin, D.H. *Int. J. Cancer* **16**, 216–225 (1975).
8. Anderson, P., Caligiuri, M., Ritz, J. & Schlossman, S.F. *Nature* **341**, 159–162 (1989).
9. Janeway, C.A. Jr *Nature* **333**, 804–806 (1988).
10. Biron, C.A., Byron, K.S. & Sullivan, J.L. *New Engl. J. Med.* **320**, 1731–1735 (1989).
11. Trinchieri, G., Santoli, D. & Koprowski, H. *J. Immun.* **120**, 1849–1856 (1978).
12. Gidlund, M. *et al. Nature* **273**, 759–761 (1978).
13. Bukowski, J.F. *et al. J. Immun.* **131**, 1531–1538 (1983).
14. Karre, K., Ljunggren, H.G., Piontek, G. & Kiessling, R. *Nature* **319**, 675–677 (1986).
15. Perussia, B. *et al. J. exp. Med.* **170**, 73–86 (1989).
16. Rembeck, R.M., Kumar, V., David, C.S. & Bennett, M. *Immunogenetics* **28**, 158–170 (1988).
17. Garcia-Penarrubia, P. *et al. J. exp. Med.* **169**, 99–113 (1989).
18. Cleviers, H., Alarcon, B., Wiseman, T. & Terhorst, C. *A. Rev. Immun.* **6**, 629–662 (1988).
19. Weissman, A.M. *et al. Science* **239**, 1018–1021 (1988).
20. Sussman, J.J. *et al. Cell* **52**, 85–95 (1988).
21. Blank, U. *et al. Nature* **337**, 187–189 (1989).
22. Brooks, C.G., Holscher, M. & Urdal, D. *J. Immun.* **135**, 1145–1152 (1985).

Topological phase of neutrons

I. J. R. Aitchison

NOVEL experimental tests of supposedly well-understood theories are always well worthwhile, even if a big surprise is unlikely. An excellent recent example¹ is the experiment performed by physicists from the Universities of Melbourne and Missouri to check the Aharonov–Casher effect², which directly tests very basic aspects of both quantum mechanics and electrodynamics. This effect is a topological quantum-interference effect, of the type first clearly identified by Yakir Aharonov and David Bohm 30 years ago³; indeed, it is closely analogous to the example these authors gave, though there are also interesting differences. One is that the effort is much weaker; thus the new observation represents an experimental *tour de force*.

Under suitable conditions, interference effects can be observed in the scattering of all wave-like disturbances, including light. The quantum wave-like aspects of matter are clearly manifested in the case of microscopic particles such as electrons or neutrons, both of which can exhibit

interference effects similar to those seen with light. The classic example is the two-slit interference effect (Fig. 1a), in which the lateral position of the characteristic fringes set up by two coherent beams depends only on their phase difference at the detector, and hence on the difference in length between the paths travelled by each beam.

The Aharonov–Bohm effect can be observed when the standard interference experiment with charged particles is modified by the introduction of a vertical, long thin solenoid between the beam paths (Fig. 1b). The solenoid produces a vertical magnetic field which, for all practical purposes, is confined inside the solenoid. Steps can be taken specifically to ensure that the electrons never penetrate the solenoid, so that the magnetic field cannot act upon them. Consequently, according to classical electrodynamics, they should experience no (Lorentz) force, and all phenomena observed should presumably be the same as those in the manifestly force-free arrangement of the unmodified experiment. Yet, according to quantum mechanics, there is an observable difference — namely, a lateral shift in the positions of the interference maxima. This is the Aharonov–Bohm effect.

Because the interference pattern is determined by the phase difference between beams following paths 1 and 2, it must be the case that the presence of the solenoid somehow affects this, even though the beams never enter the solenoid itself. To understand the quantum-mechanical interpretation of the phenomenon, one has to realize that the phase accrued by a quantum particle along a path traversed in one direction is precisely reversed in sign if the sense of traversal of the path is itself reversed. Thus the phase difference $\phi = \phi_1 - \phi_2$ between waves that traverse paths 1 and 2 in Fig. 1b can also be written as $\phi_1 + \phi_{(-2)}$ where $\phi_{(-2)}$ means that path 2 is reversed. The phase ϕ can now be interpreted as the total phase accrued on the closed path $AB_1C_1B_2A$. As Aharonov and Bohm noted, the essential point now is that the marriage of quantum mechanics with electrodynamics (called quantum electrodynamics) predicts that the phase accrued on a closed path is proportional to the value of the magnetic flux enclosed with the path — even if the path itself never actually enters the region of non-zero flux. All that matters is whether a closed path ‘links’ some magnetic flux or not.

This last point is the reason the effect is now recognized as having a distinctively ‘topological’ nature. Topology has to do with properties that remain unchanged

when shapes of things are smoothly distorted. Thus, if two circular rings are linked, they remain so as the circles are distorted into triangles, say. In the case of the Aharonov–Bohm effect, a path loop such as $AB_1C_1B_2A$ can be distorted by altering the positions of the slits B_1 and B_2 , but as long as the path still links the solenoid containing the ‘tube’ of magnetic flux, the predicted shift in the interference pattern will be unchanged.

The effect was first observed in 1960 by Robert Chambers at the University of Bristol⁴. The interpretation of both the theory and the experiment was the subject of controversy for some years, but the question has been finally settled, to most people’s satisfaction, in a recent experiment by Akira Tonomura and colleagues at the Hitachi Research Laboratories, Tokyo⁵.

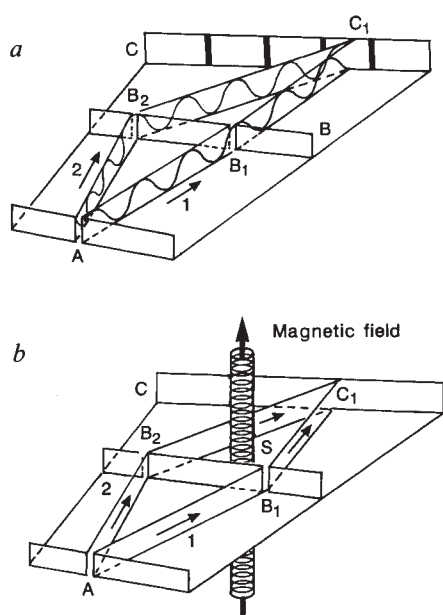


FIG. 1 a, The interference between the two components of a coherent wave split by a pair of slits (B_1 and B_2), is a classic demonstration of wave behaviour. The fringe maxima occur where the wave maxima coincide, and the minima occur where crest and trough cancel. Thus the lateral position depends only on the phase difference of each wave at the detector C and hence the difference between the number of cycles passed through on each path, 1 and 2, in traversing from source A to detector. b, The introduction of a solenoid between the two paths of charged-particle waves produces an additional phase shift — the Aharonov–Bohm effect.

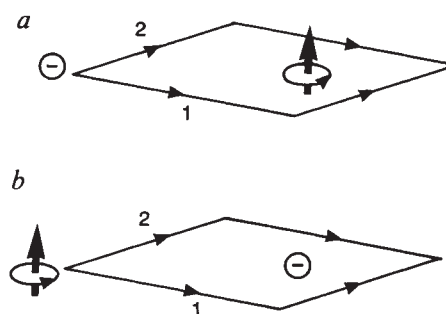


FIG. 2 The equivalence of the Aharonov–Bohm effect (a), in which charged-particle beams pass either side of a magnetic dipole, and the Aharonov–Casher effect (b), in which the beams comprise neutral magnetic dipoles, and the linked field is due to a charge.

In 1984, Aharonov and Casher² proposed an interesting variant of the Aharonov–Bohm effect. Their idea can be understood as follows. The simplest possible solenoid is one with only a single loop of wire. The magnetic field due to a small current loop is the same as that of a magnetic dipole situated at its centre and aligned with its axis. (The full solenoid is equivalent to a line of such dipoles stacked on top of each other.) The corresponding Aharonov–Bohm experiment would then consist in directing beams of charged particles on either side of a magnetic dipole and observing the resulting interference pattern (Fig. 2a). Now reverse the roles of charge and magnetic moment, so that a beam of particles possessing a magnetic moment, but no net charge, is directed on either side of a charged particle (Fig. 2b). The observation of a shift in the interference pattern relative to the case in which the enclosed charge is absent, is the Aharonov–Casher effect. Of course, the single charge in Fig. 2b can be replaced by a line of charges stacked vertically on top of each other, thus adding a further, vertical dimension to the arrangement.

The neutron is exactly such a neutral particle and, as already mentioned, neutrons can exhibit pronounced wave-like interference phenomena — which may be conveniently observed in an interferometer cut from a single crystal of silicon⁶. Thus in the new experiment of Cimmino *et al.*¹ to check the Aharonov–Casher effect, the two beam paths 1 and 2 correspond to the two ‘arms’ of a neutron interferometer. Recall that, being topological in nature, the predicted phase shift depends on neither the geometry of the beam paths nor the detailed configuration of the charge; all that matters is the linkage of the electric charge line by the loop corresponding to the beam path. Thus Cimmino *et al.* enhanced the effect by using a series of line charges — in fact, an electrode system, contained within the two beam paths, and creating a horizontal electric field (Fig. 3). The effect is manifested, in principle, via a small change in the counting rate at the point C when the

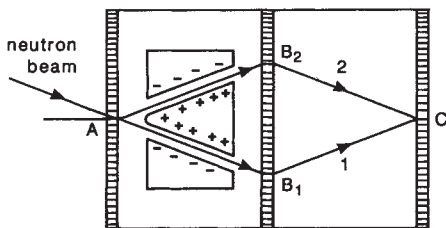


FIG. 3 Schematic of the crystalline-silicon neutron interferometer used by Cimmino *et al.* to demonstrate the Aharonov–Casher effect.

electric field is switched on, corresponding to a small shift in the intensity pattern.

A significant problem in demonstrating the Aharonov–Casher effect is that the neutron’s magnetic moment is very small, so that the phase shift is predicted to be only a few milliradians; the Aharonov–Bohm shift can easily be a thousand times larger. Furthermore, the magnetic moments of the beam neutrons should all be polarized parallel to the lines of charge. But the intensity of neutron beams is weaker than that of electron beams, and weaker still if only neutrons of a definite alignment are wanted. Very long exposure times would be needed to build up reliable statistics, but this would necessitate ensuring the mechanical rigidity of the interferometer to an extraordinary degree as the normal interference pattern is determined by the difference in path length of the two arms. In fact, Cimmino *et al.* state that an uncontrolled phase drift ϕ_0 can occur, of the order of 50 mrad a day, that would completely obscure the tiny Aharonov–Casher effect.

This problem has been overcome in a highly ingenious way. For polarized neutrons, the effect of the phase drift ϕ_0 can be cancelled by reversing the polarity of the electric field and measuring the corresponding difference in the neutron

counting rates. Unfortunately, this cancellation does not occur for an unpolarized neutron beam, but Cimmino *et al.* surmounted this by exploiting the fact that a phase shift can be induced if the two beams pass through different gravitational potentials⁷, which can easily be achieved by tilting the whole apparatus, and adjusting the phase shift, ϕ_g , so induced to cancel ϕ_0 . A further refinement was the introduction of a magnetic field over part of one of the neutron paths; this induces yet another phase shift⁸, which depends on the alignment of the neutron’s magnetic moment and maximizes the sensitivity of the experiment. The measured Aharonov–Casher phase is 1.46 ± 0.35 times the predicted value, where the error is purely statistical.

This demonstration of a subtle effect is an elegant example of the experimentalist’s art. It turns out that the detailed theoretical interpretation is not without interest either. Recall that in the Aharonov–Bohm effect, the particle beams never penetrate the solenoid and thus ‘feel’ no classical force. Do the neutrons in the new effect feel a force or not? At first sight it would seem that they do, because, according to special relativity, a particle moving with uniform velocity through an electric field also experiences an apparent magnetic field — and this can act on the neutron’s magnetic moment. Indeed, this force causes a differential slowing down between the neutrons travelling along the two paths, which generates a phase shift equal to Aharonov and Casher’s prediction⁹.

This interpretation has been challenged by Alfred Goldhaber, who points out¹⁰ that the total force on a neutron must include a contribution from the change in the momentum carried by the electromagnetic field at the position of the neutron; when this is included, the total force is indeed zero, provided the dipole moment of the neutron is perpendicular to the electric field. Thus the Aharonov–Casher effect, as one would expect from its close analogy with the Aharonov–Bohm effect, does seem to be genuinely force-free and topological. All the same, we can expect to see further discussion of this point, and further exploration of the possibilities opened up by the new effect. □

I.J.R. Aitchison is in the Department of Theoretical Physics, University of Oxford, Oxford OX1 3NP, UK.

1. Cimmino, A. *et al.* *Phys. Rev. Lett.* **63**, 380–383 (1989).
2. Aharonov, Y. & Casher, A. *Phys. Rev. Lett.* **53**, 319–321 (1984).
3. Aharonov, Y. & Bohm, D. *Phys. Rev.* **115**, 485–491 (1959).
4. Chambers, R.G. *Phys. Rev. Lett.* **5**, 3–5 (1960).
5. Tonomura, A. *et al.* *Phys. Rev. Lett.* **56**, 792–795 (1986).
6. Rauch, H. *et al.* *Phys. Lett.* **47A**, 369–371 (1974).
7. Colella, R. *et al.* *Phys. Rev. Lett.* **34**, 1472–1474 (1975).
8. Werner, S.A. *et al.* *Phys. Rev. Lett.* **35**, 1053–1055 (1975).
9. Boyer, T.H. *Phys. Rev.* **A36**, 5083–5086 (1987).
10. Goldhaber, A.S. *Phys. Rev. Lett.* **62**, 482 (1989).

Shaking fever

MANY health physicists suspect that microwaves are hazardous, even in doses too small to cause overt heating. Perhaps, says Daedalus, they set up molecular vibrations. A big protein molecule must have many fundamental vibrations at microwave frequencies. If by chance a lot of them lay in the output-band of a powerful microwave source, all those vibrational modes would be excited simultaneously.

This by itself would not be harmful. A molecule can take a lot of vibration without falling apart. But Daedalus recalls the ‘freak wave’ phenomenon: many small vibrations coinciding by chance to produce a huge and damaging momentary amplitude. If many of the molecular modes of the protein all happen to stretch a particular bond, then a freak-wave coincidence phasing them all together would probably break the bond and destroy the protein.

Daedalus sees a new therapy here. For viruses, too, are big enough to have many vibrational modes in the microwave region. But, whereas proteins come in innumerable varieties each with many molecular conformations, all the particles of a particular virus must be identical, with the same vibrational modes. So DREADCO biochemists are studying selected viruses by microwave spectroscopy, to identify every vibrational mode which contributes to stretching a particular bond or set of bonds.

The vibrational analysis of a complete virus particle will be daunting indeed; but the effort will be well worthwhile. For the DREADCO team will then be able to irradiate their viral cultures with all these crucial microwave frequencies simultaneously. They will even phase the frequencies for maximum freak-wave impact. The resulting ‘targeted microwave radiation’ will be a splendid, highly specific, antiviral therapy. Its heating effect need not be too great and, although it may zap the odd protein molecule by sheer chance, its general biological effect will be slight. But it will infallibly break up every virus particle of the targeted type.

The viral ills that flesh is heir to will thus be stopped in their tracks. For microwaves can penetrate the entire body. Even if the chosen virus hides inside the cells of its victim, where antibodies cannot reach, it will be swiftly shattered by microwave-induced freak-wave scission. Once the vibrational analysis of a given virus has been completed, DREADCO’s programmable multifrequency microwave generator will be able to blast it out of the body on demand. Influenza, glandular fever, hepatitis, even the common cold, will surely succumb to this potent therapy. Even if the virus mutates in self-defence, a re-analysis of its new form will put it firmly back in DREADCO’s sights. David Jones

Bamboo flowering and pandas

SIR—The flowering and subsequent natural dieback of arrow bamboo in Wolong and other areas in China in 1983 did not cause the mass starvation of pandas that was feared because they foraged on alternative bamboo species or unflowered arrow bamboo patches¹⁻³. Now that the perceived crisis has passed, efforts to conserve the panda can focus on the urgent problems of poaching, habitat preservation, release of captive animals and development of a captive-breeding programme^{1,4,5}. But an unusual opportunity to restore degraded panda habitat should receive equal attention.

Even though bamboo density is greater in those areas of Wolong that were clear-felled in the 1960s and 1970s than in the still-forested areas, pandas prefer forest habitat. Few trees, especially conifers, became re-established in the clearcuts because of the dense vegetatively propagated bamboo, but conifers dominate the subalpine forest in Wolong and old hollow conifers serve as maternity dens for pandas⁵⁻⁷.

We therefore urge the immediate planting of native conifers in the areas cleared by the 1983 dieback before the regrowing arrow bamboo has reached a size that would inhibit conifer establishment — this

requires 10–15 years. Rapid development of conifer forest from planted seedlings may also help reduce the high bamboo seedling mortality observed in clearcuts⁸. Cost-effective opportunities to restore degraded panda habitat to conifer forest are infrequent because arrow bamboo will not flower again for 45–50 years^{5,9}.

ALAN H. TAYLOR

Department of Geography,
University of Maryland Baltimore County,
Baltimore, Maryland 21228, USA

QIN ZISHENG

Department of Biology,
Nanchong Teacher's College,
Nanchong, Sichuan,
People's Republic of China

- Schaller, G. B. *Nature* **327**, 562 (1987).
- Taylor, A. H. & Qin Zisheng *J. appl. Ecol.* **24**, 419–433 (1987).
- Johnson, K. G., Schaller, G. B. & Hu Jinchu *Nat. geog. Res.* **4**, 161–177 (1988).
- O'Brien, S. J. & Knight, J. A. *Nature* **325**, 758–759 (1987).
- Schaller, G. B., Hu Jinchu, Pan Wenshi & Zhu Jing *The Giant Pandas of Wolong* xiii, 298 (University of Chicago Press, 1985).
- Taylor, A. H. & Qin Zisheng *J. Ecol.* **76**, 1204–1218 (1988).
- Taylor, A. H. & Qin Zisheng *Biol. Conserv.* **47**, 83–108 (1989).
- Taylor, A. H. & Qin Zisheng *Am. J. Bot.* **75**, 1065–1073 (1988).
- Campbell, J. J. N. & Qin Zisheng *J. Bamboo Soc. Am.* **4**, 1–34 (1983).

Basis of hepatitis delta virus disease?

SIR—Symons¹ has recently discussed a reported² sequence complementarity between the six viroid 'species' that cause disease in tomato plants and tomato 7S RNA, a small cytoplasmic RNA present in various organisms as a component of the signal recognition particle, a structure involved in the translocation of secretory and membrane-associated proteins. It has been suggested that formation of hybrids between viroids and 7S RNA may be a

RNA replication intermediate³. Also, limited sequence similarity has been reported between viroids and HDV RNA⁴.

We have found that an extensive complementarity exists between two regions of human 7SL RNA⁵ (the counterpart of tomato 7S RNA) and anti-genomic HDV RNA⁶ (see figure). The first region involves nucleotides 10–55 of 7S RNA and 683–724 of the HDV replicative intermediate. In a 48-nucleotide domain, 35

```

7SL 10  GGUGGCGCGUGCCUGUAGUCCAGC-UACUCGGGAGGC-UGAGGCGUG 55
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
HDV 683 CUACCG-GC-CGUAC--CAGGGUGCGAGGAGC-GACCGCG-GCCGACC 724
7SL 188 GCCCAGGUGCGGAACG-GAGCAGGUCAAACUCCCGUGCUGAUC 230
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
HDV 858 CGGGUCCAGCCUG-GCGCUCCUCCACCUCUACGG-UACGGCGUG 899

```

Proposed base-pairing between human 7SL RNA and anti-genomic HDV RNA. Colons indicate G:U base-pairs. Lowercase letters denote the two nucleotides that are identical in two reported HDV RNA sequences^{4,6} but differ in a third⁷.

factor in disease^{1,2}. We wish to extend that suggestion to the hepatitis delta virus (HDV), a human pathogen that causes severe liver disease (delta hepatitis) throughout the world³.

Similarities between viroids and HDV have been demonstrated previously. Like viroids, HDV has a circular, single-stranded RNA genome that possesses a high degree of self-complementarity⁴. HDV RNA is believed to replicate via a rolling cycle mechanism through an anti-genomic

(73%) bases are complementary (including two G:U base-pairs); three stretches of 6–8 nucleotides are perfectly base-paired. The second region of complementarity is found between bases

188–230 of 7SL RNA and 858–899 of anti-genomic HDV RNA. In a 44-nucleotide domain, 34 (77%) bases are complementary (including three G:U base-pairs). Two long stretches, 12 and 9 bases in length, of perfectly paired nucleotides are found at the ends of this domain. The sequences that separate these two regions, although approximately of the same length (132 and 133 nucleotides), do not show any significant complementarity. Of the 69 nucleotides of the anti-genomic

HDV RNA complementary to 7SL RNA, 67 are invariant among the three known HDV sequences^{4,6,7}, whereas other regions of HDV show a rather high heterogeneity. Thus, the region of HDV anti-genome exhibiting complementarity with 7SL RNA is highly conserved among the known virus isolates.

The suggested base-pairing between human 7SL RNA and anti-genomic HDV RNA is even more extensive (and possibly more stable) than that between tomato 7S RNA and the six viroids. Interestingly, the region in tomato 7S RNA (bases 26–67) thought to be capable of annealing to viroid RNA corresponds to one of the regions of the human 7SL RNA complementary to the anti-genomic HDV RNA. However, due to sequence differences between tomato 7S RNA and its human counterpart², there is no significant similarity between anti-genomic HDV RNA and viroid RNA in this region.

Thus, just as viroids may cause disease by base-pairing to a host structural RNA, so may a replicative intermediate of HDV. As HDV is known to replicate inside the nuclei of infected hepatocytes⁸, its anti-genomic form may anneal to the 7SL RNA shortly after the host RNA is transcribed and before it becomes incorporated in the signal recognition particle. This hypothesis is consistent with the notion that HDV is directly cytopathic³ and can be tested experimentally. If true, it might lead to new therapeutic strategies for the control of delta hepatitis.

FRANCESCO NEGRO

JOHN L. GERIN

Division of Molecular Virology and
Immunology,
Georgetown University,
Rockville, Maryland 20852, USA

ROBERT H. PURCELL

ROGER H. MILLER

Hepatitis Viruses Section,
National Institute of Allergy and
Infectious Diseases,
Bethesda, Maryland 20892, USA

- Symons, B. *Nature* **336**, 542–543 (1989).
- Haas, B. *et al. EMBO J.* **7**, 4063–4074 (1988).
- Rizzetto, M. *et al.* (eds) *Hepatitis Delta Virus and its Infection* (Liss, New York, 1987).
- Wang, W.S. *et al. Nature* **323**, 508–514 (1986).
- Ullu, E. & Tschudi, C. *Nature* **312**, 171–172 (1984).
- Makino, S. *et al. Nature* **329**, 343–346 (1987).
- Kuo, M.Y.P. *et al. J. Virol.* **62**, 1855–1861 (1988).
- Gowans, E.J. *et al. Virology* **167**, 274–278 (1988).

Aurorae aplenty

SIR—Schröder's suggestion¹, that the Maunder minimum² of solar activity does not show up in European auroral records, is corroborated by historical evidence from the Soviet Union. Although the great aurora of 16 March 1716 (Julian) seemed to awaken the scientific society of the time, initiating studies of the ionosphere, there were nevertheless useful observations made in the middle latitudes

of Eurasia for 70 years before this.

My collection includes observations from Russian, Ukrainian, Polish and Armenian sources, as well as some German ones and an English text recounting that Patrick Gordon, a general and rear-admiral in the service of Peter I, saw the great aurora in Kiev on 1 June 1685. Only a fraction of these sources are known to historians of geophysics, much of the information having been derived by Yu. A. Mytsyk, of Dnepropetrovsk University, and myself from antique publications and manuscripts.

Scientific observations in Russia date back to 1725, but the tradition of chronicle-writing seems to have failed in the mid seventeenth century. But in that period several new sources arose: Ukrainian chronicles, which appeared at a time of national revival; Siberian annals owing their origin to a rapid Czarist expansion to the East and the foundation of autonomous Siberian cultural centres such as Tobolsk and Tomsk; and new Russian translations of West European records, the originals of which may be lost.

This last category includes the early periodicals (*Zeitung*, *Anzeigen*, *Berichten* and so on). Muscovite commercial and diplomatic agents gathered information from these sources and sent it to Russia. A single hand-written copy of *Vesti-Kuranty*, compiled at the Czar's court, has been preserved in Russian archives, recording German, Danish, Swiss and Hungarian aurorae otherwise unknown to West European scientists.

Russian sources record 57 (perhaps 61) auroral phenomena from 1645 to 1715. Schröder¹ thinks there were 110 (perhaps 130) cases in Central Europe, and the late Professor M. Keimatsu's collections contain 26 observations from the Far East. All in all there are 193 (perhaps 217) cases, eight of which coincide (for example, one aurora was seen simultaneously in Chernigov and Kyoto).

During the 70 years, about 185 (209?) aurorae were noted, an average rate of 2.8 (3.0?) a year. This is similar to the typical rate of aurorae now seen in Moscow or Berlin. We can conclude that the term Maunder minimum should be placed between inverted commas when it refers to auroral activity.

VICTOR A. LOISHA

Internationalists-Street 22,
51 Tomsk, Siberia, 634060 USSR

1. Schröder, W. *Nature* **335**, 676 (1988).
2. Eddy, J.A. *Science* **192**, 1189 (1972).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of less than 500 words and five references. □

Lateral preference and longevity

SIR—Longevities of right-handed and left-handed baseball players have been compared^{1,2} unadjusted for birth date. To obtain a potentially more accurate paired comparison, I treat each birth year as a cohort. I have included players born up to 1922 classified by throwing hand (as in ref. 2) and use 'R-L' to designate the mean longevity of right-handers minus that of left-handers within each birth year.

R-L has a higher variance towards the earliest and latest birth years because these contain the smallest proportion of the 4,479 deceased players³ studied. To obtain a homogeneous variance as assumed in regression, successive R-L estimates were combined into groups, beginning at the first year, until each group included at least 25 left-handers. (Left-handers comprise 18 per cent of the player population and therefore contribute most of the variance.) The resultant 23 groups each included 25 to 50 left-handers. The mean birth date and R-L for each group were the averages weighted by the number of left-handers in each birth year.

Using this grouped data, the mean R-L is zero but there is a significant ($P < 0.01$) negative linear regression of R-L on birth date, B : $R-L = 3.36 - 1.38 B$ where B is (birth year - 1868.54)/10.

Obviously, R-L cannot decline indefinitely. It is plausible that stress factors¹ or discrimination against sinistrals decreased left-handers' longevity historically. Our lifespan has presumably been genetically selected for good reproductive fitness. There would, therefore, have been selective pressure to evolve a linkage between left-handedness and a factor for increased longevity, partially offsetting the lower longevity otherwise associated with left-handedness.

If such linkage exists, removal of handedness-related stresses in modern times would have shifted R-L from positive to negative, just as the preceding regression equation implies. According to this hypothesis, R-L would approach a negative value asymptotically as handedness-related stresses decrease at various rates. Therefore, an asymptotic regression⁴ was fitted:

$$R-L = -3.73 + \frac{8.69}{(s.e. = 1.22)} \times 0.67^x$$

$$(P < 0.01) \quad (P < 0.01)$$

Right-handers born before 1890 had an advantage. Thereafter, left-handers had an advantage asymptotically approaching 3.73 years. Without extrapolating, left-handers' advantage was estimated at 2.1 years for those born in 1910, with 95 per cent confidence limits of 0.4 to 3.8 years. Presumably this relationship holds

generally, rather than resulting from the players-selection process or a fluke. The advantage of being left-handed parallels the advantage of being female in affluent societies, and I suggest these have a similar evolutionary explanation. According to my hypothesis, it is predicted that left-handers will be found to have greater longevity than right-handers among Americans on average; conversely, sinistrals' longevity will be lesser in countries with low and approximately equal male and female longevity still prevailing.

MAX G. ANDERSON

Canadian Statistical Analysis Service,
2942 West 41st Avenue,
Vancouver, BC, V6N 3C7, Canada

1. Halpern, D.F. & Coren, S. *Nature* **333**, 213 (1988).
2. Wood, E.K. *Nature* **335**, 212 (1988).
3. Neft, D.S. & Cohen, R.M. *The Sports Encyclopedia: Baseball* 8th edn (St Martin's Press, New York, 1988).
4. Snedecor, G.W. & Cochran, W.G. *Statistical Methods* (Iowa State University Press, Ames, 1967).

Dating of the Ugarit eclipse

SIR—De Jong and van Soldt¹ have re-analysed clay tablet KTU 1.78 found in Ugarit in 1948 and suggested that the text on the tablet pertains to a solar eclipse which occurred in the second millennium BC. As Walker² has already pointed out, their dating is speculative. In addition to Walker's Assyriological reservations, an objection on astronomical grounds is also necessary.

The text refers to a phenomenon which may very well be interpreted as a solar eclipse: (Obverse) "On the ... day of the new moon in (the month) hiyaru the Sun went down, its gate-keeper was Ršp"; (Reverse) "Two livers were examined: danger". Tacitly de Jong and van Soldt assume that the eclipse was total, though this is by no means clear from the text. The Ugaritic terminology is not sufficiently well known to identify the term "the Sun went down" as a total eclipse; it might have been a partial eclipse. The 'gate-keeper' Mars in the text does not help to clarify the situation either, because its occurrence might just be the result of an astrological calculation or of a connection of the planet with a certain month or year. It is interesting to note in this respect that the only preserved material related to tablet KTU 1.78 consists of two fragments of almanacs and one astrological text.

The last possible reason for considering the eclipse text to refer to a total eclipse is its mere survival. The assumption is often made that when only a few records of solar

1. De Jong, T. & van Soldt, W.H. *Nature* **338**, 238-240 (1989).
2. Walker, C.B.F. *Nature* **338**, 204-205 (1989).
3. Newton, R.R. *The Moon's Acceleration and its Physical Origins*, vol 1. (John Hopkins University Press, Baltimore, 1979).
4. Bruch, H. *Chronographia Johannis de Beke* (Martinus Nijhoff, Gravenhage, 1973).
5. Mostert, R. *Quart. J. Roy. astr. Soc.* **30**, 113-114 (1989).

eclipses survive, they refer of necessity to total eclipses. This is not necessarily true, as was already pointed out by Newton³. Furthermore, I recently presented a solar eclipse record which can serve as a counter-example, namely the only account of a solar eclipse from the *Chronographia*, a fourteenth century historical work. It is the only record of a solar eclipse in that work, but nevertheless it is with certainty a record of a partial eclipse^{4,5}. It is there-

fore possible that the Ugarit record also refers to a partial eclipse. If so, dating the record is not feasible in view of the larger number of possibilities. In conclusion, the record of a solar eclipse of the second millennium is still not dated with certainty and conclusions drawn from the dating should be met with some reservation.

R. MOSTERT

Vondelstraat 270,
1814 MC Alkmaar, Netherlands

Epithelial homing of $\gamma\delta$ T cells?

SIR — Only one or two per cent of T lymphocytes in systemic lymphoid organs or blood, of mice and men, express the $\gamma\delta$ T-cell receptor (TCR) for antigen; this minor fraction mainly consists of cells that carry the CD3 antigen but not the CD4 or CD8 antigens and so are termed CD3⁺CD4⁻CD8⁻ or 'double-negative'¹. It was therefore remarkable when intraepithelial lymphocytes (IEL) isolated from the murine gut were reported to be exclusively CD3⁺CD4⁻CD8⁺ and, by immunoprecipitation, seemed to bear mainly $\gamma\delta$ TCR (refs 2 and 3). Similar results were obtained by immunohistochemistry for avian and apparently also for human IEL⁴. A tissue-specific distribution of TCR subtypes has now been proposed⁵, but has received a sceptical response^{6,7}.

We tested this theory with BMA031, an $\alpha\beta$ framework antibody⁸, and 11F2 that reacts with all types of $\gamma\delta$ TCR (ref. 9). In immunocytochemical tests on a CD3⁺ double negative $\gamma\delta$ T-cell line, these monoclonal antibodies behaved correctly; but they performed unsatisfactorily on cryostat sections. Labelling of viable

T cells was therefore performed by permeation of fresh jejunal biopsy specimens with these two antibodies and with antibodies to CD8 and CD4 (Dakopatts)¹⁰. Membrane expression of $\alpha\beta$ TCR was seen abundantly in the lamina propria and also on a variable number of IEL (Fig. 1a). The proportions of CD8⁺ and $\alpha\beta$ TCR IEL obtained in the seven normal cases were well correlated (Pearson's $r=0.865$). To compensate for variable antibody diffusion, the results for $\alpha\beta$ TCR, $\gamma\delta$ TCR, and CD4⁺ IEL were adjusted by a factor derived from the corresponding CD8⁺ to CD3⁺ ratio, taking 90% as the average 'true' value¹¹. The median $\alpha\beta$ TCR fraction of CD3⁺ IEL was thus estimated to be 92% (64–100%, $n=3$) with labelling at 37 °C for two hours and 70% (41–100%, $n=4$) at 4 °C for two to eight hours. Conversely, the proportion of $\gamma\delta$ TCR was usually below 8% and that of CD4⁺ IEL below 13% (median, 9%).

The figures obtained for three coeliac disease specimens indicated that there might be a relative increase of $\gamma\delta$ TCR (1–16%) compared with the $\alpha\beta$ TCR

(median, 48%) and CD4⁺ (median, 9%) IEL. Such a disease-associated change (Fig. 1b) could reflect the increased intraepithelial CD3⁺ double-negative subset recently identified in this disease¹².

In conclusion, we have shown that most human intestinal IEL express the $\alpha\beta$ TCR. Moreover, almost half of the CD3⁺ IEL were found to be positive for the UCHL1/CD45 molecule¹⁰; this implies that they are probably CD3⁺CD8⁺ $\alpha\beta$ memory T cells.

We have recently used another set of monoclonal antibodies, namely β F1 and TCR δ 1 from T Cell Sciences (Cambridge, Mass.), and have confirmed by immunohistochemistry on cryostat sections that $\alpha\beta$ TCR cells are strikingly predominant among human IEL, and that the $\gamma\delta$ TCR fraction is indeed increased (from a normal median of 2% to 22% in treated or untreated coeliac disease). The CD8 expression of this fraction is decreased (from 25% to 10%) and CD4 expression is not seen. Also notable is the finding that the intestinal $\gamma\delta$ TCR⁺ cells are mainly located in the epithelium, both normally and in coeliac disease (T.S. Halstensen *et al.*, *Scand. J. Immunol.*; in the press).

PER BRANDTZAEG

TROND S. HALSTENSEN

HELGE SCOTT

LUDVIG M. SOLLID

KOLBJØRN VALNES

Laboratory for Immunohistochemistry

and Immunopathology (LIIPAT),

Institute of Pathology,

University of Oslo,

Rikshospitalet, Oslo, Norway

VIDAR BOSNES

Institute of Transplantation Immunology,

University of Oslo,

Rikshospitalet, Oslo, Norway

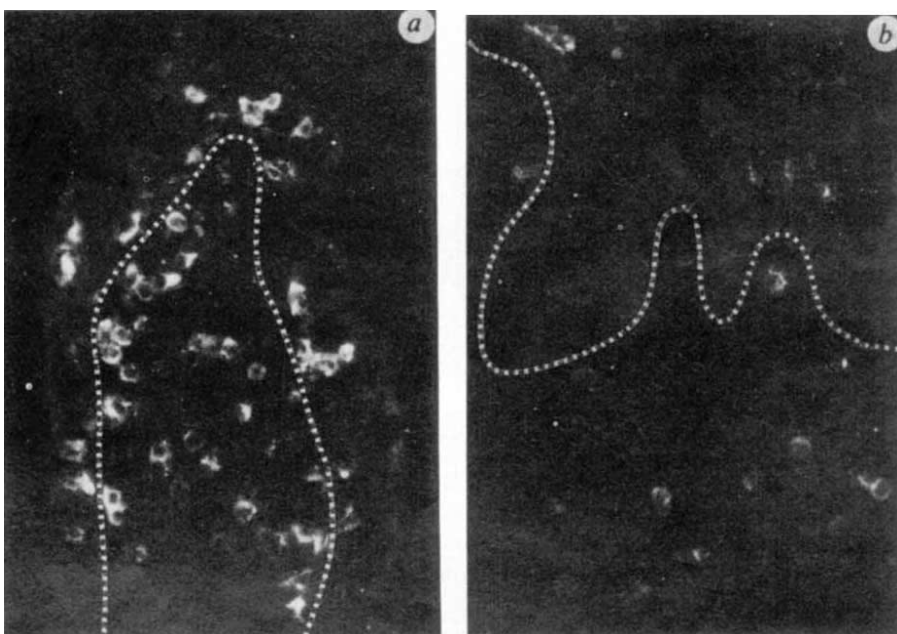


FIG. 1 a, Immunofluorescence staining of $\alpha\beta$ TCR in normal jejunal mucosa, and b, of $\gamma\delta$ TCR in coeliac disease mucosa with villous atrophy after permeation of the tissue specimens with monoclonal antibodies BMA031 (Behring, Marburg, FRG) and 11F2 (J. Borst), respectively. Mucosal surface at the top. Epithelial basement membrane indicated by broken line (x400).

SIR—Regarding the controversy that has already surfaced in *Scientific Correspondence* on the frequency of $\gamma\delta$ T cells in human intestinal epithelium^{7,13} my collaborators, Chen-lo Chen and Max Cooper, and I have investigated the tissue localization of cells expressing the δ -chain with the monoclonal antibody to TCR δ 1 of Michael Brenner. This appears to recognize a constant region of the TCR δ chain¹⁴, unlike δ TCS1 used by Spencer and Isaacson⁷, which recognizes a subset of the $\gamma\delta$ T cells¹⁴. We found that 80% of $\gamma\delta$ T cells in the intestinal mucosa are localized in the epithelium while the remaining 20% are in the lamina propria¹⁵. Nevertheless, these cells comprise a minor subset (about 7%) of the intraepithelial CD3⁺ cells. In the chicken⁴, we observed the same $\gamma\delta$ T cells homing into the intestinal epithelium but it is noteworthy that only half of the epithelial CD3⁺ cells express the $\gamma\delta$ TCR. A majority of these are CD8⁺, and none express CD4.

The preferential tissue localization

pattern of $\gamma\delta$ T cells must be an important clue to the function of this T cell sublineage¹⁶. However, the distinctive tissue localization pattern is not strictly an epithelial tropism. The preferential homing of $\gamma\delta$ T cells to the epidermis that is seen in the mouse, is not observed in either chickens or humans. Furthermore, we have observed a selective homing pattern in the spleen that is phylogenetically conserved. In both chickens and humans^{4,15} the $\gamma\delta$ T cells preferentially localize in the splenic red pulp rather than the periarteriolar lymphoid sheath, which contains the majority of the splenic $\alpha\beta$ T cells. Since the relationship of the splenic red pulp to a system of epithelial defence is not obvious, these observations suggest a more general function for $\gamma\delta$ T cells rather than a defence role restricted to the epithelial boundaries of the body.

R. PAT BUCY

Department of Pathology,
University of Alabama at Birmingham,
Birmingham, Alabama 35294, USA

SIR—We would like to contribute to the discussion concerning the preferential localization of $\gamma\delta$ T cells in human epithelia. In the murine system it has been observed that $\gamma\delta$ T cells are present in epidermis and intestinal epithelium in much higher frequency than $\alpha\beta$ T cells. Janeway in particular, has elaborated on this finding and put forward the general suggestion that $\gamma\delta$ T cells could play a particular role in immunosurveillance at these sites^{5,16}.

We have looked into the distribution of $\gamma\delta$ T cells versus $\alpha\beta$ T cells in a variety of normal human tissues of lymphoid and non-lymphoid nature. Serial sections of freshly frozen tissue were acetone fixed, incubated with anti-CD3, anti-TCR $\gamma\delta$ -1 (ref. 9) and β F1 (anti-TCR $\alpha\beta$) (ref. 17) monoclonal antibodies and stained according to a highly sensitive immuno-alkaline phosphatase method. We studied bone marrow; fetal and neonatal liver; fetal and neonatal thymus; fetal, neonatal and adult lymph nodes; adult spleen and tonsil; neonatal and adult skin^{18,21}, oesophagus, stomach, small and large intestine, lung and trachea, testis, epididymis, uterus, vagina, kidney and brain.

At all sites within these organs $\alpha\beta$ T cells constituted over ninety per cent of CD3⁺ cells. $\gamma\delta$ T cells comprised less than ten per cent of CD3⁺ cells. Therefore, we conclude that there is no evidence for preferential homing or localization of $\gamma\delta$ T cells in human epithelia. However, in none of the organs were $\gamma\delta$ T cells observed in lymph follicles (of lymph nodes, tonsil, spleen, Peyer's patches or lung). Also, the percentage of $\gamma\delta$ T cells in epidermis seemed on average higher than in papillary dermis. Therefore, there may still be a difference in migration of

$\gamma\delta$ versus $\alpha\beta$ T cells, but this does not result in their preferential localization in epithelia. With respect to morphology of $\gamma\delta$ T cells and $\alpha\beta$ T cells, we find that generally they cannot be distinguished, but occasional CD3⁺ cells with dendritic appearance always proved to express $\gamma\delta$ TCR.

In view of these findings, great caution should be taken in ascribing a special function to $\gamma\delta$ T cells in immunosurveillance of human epithelia. Future intensive and thorough investigation should reveal what contribution these cells make to the immune system.

THEA M. VROOM

Daniel den Hoed Cancer Centre,
Rotterdam,

JAN D. BOS

Academic Medical Centre,
University of Amsterdam,

JANNIE BORST

The Netherlands Cancer Institute,
1066CX Amsterdam,
The Netherlands

JANEWAY REPLIES—The above letters raise two questions. First, do T cells bearing $\gamma\delta$ TCR have a special tropism for epithelia, and second, is this tropism seen in man? In addition, the letters appear to contain a misapprehension of the content of two articles^{5,16} that I have written on this topic. These articles propose that one role $\gamma\delta$ T cells might have is defence of epithelia from intracellular infection and/or transformation, and that such cells might recognize not foreign antigen but an alteration in the host cell surface induced by infection. Both articles clearly state that such T cells are also found in other sites where they must mediate other functions. Thus, I do not take issue with the conclusion of the above letters that epithelial defence is not the only role such cells play. There are, however, strong data arguing that some T cells of this class exhibit this specialization.

Since this proposal was made, we and others have shown that T cells bearing $\gamma\delta$ TCR are in fact the dominant population in murine intestine^{2,3}. More importantly, we have shown that such cells express receptors encoded by the $V_{\gamma 7}$ gene V_{γ} not expressed by $\gamma\delta$ T cells in any other tissue^{3,19}. S. Tonegawa *et al.* have

shown that the $V_{\gamma 6}$ gene, again not expressed by $\gamma\delta$ T cells in other tissues, is used in reproductive epithelium (personal communication), while Allison *et al.*²⁰ have shown that the $V_{\gamma 5}$ gene is expressed only by $\gamma\delta$ T cells in epidermis. Thus, in addition to tropism to epithelia, there is a marked restriction in the specificity of $\gamma\delta$ T cells found in these tissues. This does not prove that T cells with $\gamma\delta$ TCR are specialized for epithelial defence, but it certainly is most readily explained in the context of our earlier hypothesis. Recently, experiments with mice transgenic for $\gamma\delta$ TCR show that epithelial homing is a property of the $\gamma\delta$ T cell lineage and not of particular $\gamma\delta$ TCR specificities. Specificity appears to reflect ligand recognition in the epithelia (M. Bonneville *et al.* personal communication). This further supports a functional role for those $\gamma\delta$ T cells in epithelia. Perhaps studies of V-gene usage in $\gamma\delta$ TCR in humans would show similarly specific localization.

The human studies cited show predominance of T cells with $\alpha\beta$ TCR in intestinal epithelium. From examining photographs of such sections, however, Bucy notes (above) that the T cells with $\gamma\delta$ TCR show tropism for the epithelial layer whereas T cells with $\alpha\beta$ TCR do not. Thus, the human data on tropism of $\gamma\delta$ TCR-bearing cells to epithelia do not conflict with those in the mouse. Rather, this tropism is masked by the presence of large numbers of T cells with $\alpha\beta$ TCR found in humans but not mice or chickens, where similar studies show 20–40% of T cells with $\alpha\beta$ TCR. Thus, the obvious species difference may reflect differences in the behaviour of $\alpha\beta$ cells and/or differences in the actual $\gamma\delta$ receptors, rather than differences in function or tropism of T cells with $\gamma\delta$ TCR. Until these issues are resolved, it is as unwise to state that T cells with $\gamma\delta$ TCR are *not* specialized for epithelial immunity as it is to state that this is their only function, or that only $\gamma\delta$ T cells can perform such functions, neither of which I have claimed (although such claims have been attributed to me).

CHARLES A. JANEWAY JR

Department of Immunology,
Howard Hughes Medical Institute,
Yale University School of Medicine,
New Haven, Connecticut 06510, USA

- Moretta, L. *et al.* *Clin. Immunol. Immunopathol.* **50**, S117–S123 (1989).
- Goodman, T. & Lefrançois, L. *Nature* **333**, 855–858 (1988).
- Bonneville, M. *et al.* *Nature* **336**, 479–481 (1988).
- Bucy, R. P., Chen, C.-L. H., Cihak, J., Löscher, U. & Cooper, M. D. *J. Immunol.* **141**, 2200–2205 (1988).
- Janeway, C. A., Jones, B. & Hayday, A. *Immun. Today* **9**, 73–76 (1988).
- Carter, L., DenHollander, N., Ramaswamy, K., Riley, D. & Befus, D. *Immun. Today* **9**, 368–369 (1988).
- Spencer, J. & Isaacson, P. G. *Nature* **337**, 416 (1989).
- Lanier, L. L., Ruitenberg, J. J., Allison, J. P. & Weiss, A. in *Leucocyte Typing III. White cell differentiation antigens*. (McMichael, A. J. *et al.* eds) (Oxford University Press, 1987).
- Borst, J. *et al.* *J. exp. Med.* **167**, 1625–1644 (1988).
- Brandtzaeg, P. *et al.* *Scand. J. Immunol.* **30**, 123–128 (1989).
- Brandtzaeg, P. *et al.* *Gut* **29**, 1116–1130 (1988).
- Spencer, J. *et al.* *Gut* **30**, 339–346 (1989).
- Spencer, J., Isaacson, P. G. & Janeway, C. A. *Nature* **337**, 416 (1989).
- Faure, F. *et al.* *J. Immunol.* **141**, 3357 (1988).
- Bucy, R. P., Chen, C.-L. H. & Cooper, M. D. *J. Immunol.* **142**, 3045–3049 (1989).
- Janeway, C. A. *Nature* **333**, 804–806 (1988).
- Brenner, M. B. *et al.* *J. Immunol.* **138**, 1502–1509 (1987).
- Bos, J. D. *et al.* *J. invest. Derm.* **92**, 406 (1989).
- Kyes, S., Carew, E., Carding, S., Janeway, C. A. & Hayday, A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5527 (1989).
- Arsanow, D. M. *et al.* *Cell* **55**, 837 (1988).
- Bos, J. D. *et al.* *J. invest. Derm.* (in the press).

Making a killing

Christopher C. Booth

Health for Sale: Quackery in England 1660–1850. By Roy Porter. Manchester University Press: 1989. Pp.280. £19.95, \$35.

HOGARTH'S famous engraving "The Company of Undertakers" portrays a group of pompously bewigged physicians together with a trio of famous quacks. There was, for the eighteenth-century patient, not much to choose between them. To little avail, the doctors bled, purged, blistered and clystered, cloaking their ignorance behind the Latin that was supposed to be the mark of their erudition. They did not always get away with it. William Cobbett described the Philadelphia physician, Benjamin Rush, known in his day as the American Hippocrates, as "a poisonous transatlantic quack", a "Samson of medicine who slaughtered thousands of citizens". But for the most part the conventional profession was regarded as the fount of medical knowledge, and all others pretending to benefit the public with their nostrums were considered to be quacks, charlatans or mountebanks.

There was, however, not a great deal of difference between the recipes of quacks and those of the established profession. Until as late as 1786, the London physician could officially prescribe the Venice treacle or Mithradatum, a remedy that contained as many as 65 ingredients including the dried flesh of vipers. Furthermore, no physician was reluctant to usurp the therapies used by herb-women and others, who were not considered by the suffering humanity of two centuries ago to be the artful, tricking practitioners of physic that Samuel Johnson regarded with such malevolence. The Earl of Chesterfield was as likely to sample the nostrum of a quack as to consult the owner of a gold-headed cane. We also do well to remember that digitalis was discovered in a family recipe for the dropsy kept by an old herb-woman in Shropshire, and that the development of both ophthalmic surgery and of orthopaedics owes much to the irregular practitioner.

In *Health for Sale*, Roy Porter sets out to put quackery in England in its historical context between 1660 and 1850. He begins with an interpretation of quackery, emphasizing how medical history must be seen as part of economic and social history as a whole, how sufferers interrelated with

the ambitions of practitioners, both regular and irregular, and how the practices of quacks cannot be divorced from a consideration of the orthodox. He sees the quack, no less than the regular doctor, as an entrepreneur in a consumer society, serving a public whose desire for medicines was only matched by the enthusiasm with which practitioners of all types dispensed

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Good vibrations — Mesmer's tub, an etching from 1790. Rods connected to a 'magic box' allegedly provided healing through transfer of vibrational energy.

them. He tells us of the career of quackery, how it ranged from the anonymous itinerant to the most famous in the metropolis — Mrs Mapp, for example, the bone-setter of Epsom, who was portrayed in Hogarth's print and who rode to town in style in her coach and four.

Perhaps the most remarkable of them all was Mrs Joanna Stephens, whose medicine for the treatment of the stone had reputedly relieved Sir Robert Walpole. Parliament rewarded her with the astounding sum of £5,000 for revealing her secret remedy, which was found to consist of "eggshells and snail shells, both calcined, wild carrot seeds, burdock seeds, ashen keys, hips and haws, all burnt to blackness and mixed with alicant soap and honey". It seems to have been remarkably ineffective, since there were still a number of stones in Sir Robert's bladder at his death. Yet Mrs Stephen's lithotryptic had its influence on the history of science, for it led the remarkable Reverend Stephen Hales to suggest that its active ingredient was lime; it was lime water that led Joseph Black to the experiments in which he discovered carbon dioxide.

With the breadth of scholarship that is his characteristic, Porter deals with the culture of quackery, with advertising, and with images of health, disease and cure. Quacks were very much involved in sexual

matters, particularly in the treatment of venereal disease. There were, furthermore, quacks who were not only peddling pills. The more famous, such as Cagliostro and John (Chevalier) Taylor, were not averse to parading their own prowess as seducers in purveying their wares. James Graham's famous Celestial Bed in his Temple of Hymen at the Adelphi, where childless couples could seek success surrounded by currents of electricity, was the *chef d'oeuvre* of a "wizard showman flourishing in the lubricious twilight-world of sex aids".

Quacks frequently became embroiled in quarrels, but their differences with the doctors almost always worked to their own advantage. Of all those of the regular profession who sought to dispose of the quacks, however, there was none so vitriolic nor so successful as Thomas Wakley, who in his newly founded journal, *The Lancet*, attacked them with particular venom and hatred. Quackery was now on the retreat, for the nineteenth century was to be a period when, with the emergence of regular, professional medicine, it gave way to a popular fringe, often motivated less by the pursuit of filthy

lucre than by the credo of the crusader.

Porter gives a highly readable account of his subject, and he is at his surest when dealing with what he describes throughout as "the long 18th century". All the well-known and many of the lesser lights are here. His view of medical history embraces not only the lives of the practitioners he so sympathetically describes, but also the views of their long-suffering patients. The work is admirably referenced and there is an excellent bibliography.

There is, however, a tendency to unnecessary repetition which could have been avoided by more careful editing. There may also be those who would have preferred a more chronological account, with perhaps some analysis of the nature of illness in England at the beginning and end of the period under review. After all, in 1660, the nation still had to experience the horrors of the last great plague in England, an epidemic which played an important role in the relationship between regular and irregular practitioners. By 1850 the pattern of infectious disease had changed. With the increased rapidity of travel, cholera had reached these shores. Diseases, no less than those who seek to treat them, have a history of their own. □

Sir Christopher Booth, formerly Director of the MRC Clinical Research Centre, is Harveian Librarian of the Royal College of Physicians, 11 St Andrews Place, London NW1 4LE, UK.

Inside story

Peter J. Smith

Theory of the Earth. By Don L. Anderson. Blackwell Scientific: 1989. Pp. 366. £49.50, \$65.

GEOPHYSICS and geochemistry each has its own special techniques and a huge body of accumulated knowledge. By definition they are distinct disciplines and have, by tradition if not necessity, been studied by distinct groups of people who have communicated with each other all too infrequently. Indeed, such are the pressures to specialize that disciplinary segmentation has gone even further, with Earth scientists now describing themselves solely as seismologists, geomagnetists, geodesists, petrologists, mineralogists, meteorite specialists, or any one of a number of other '-ists'.

Unfortunately, the Earth's interior not only fails to acknowledge such divisions, it insists on dragging in a whole range of topics not recognizably geoscientific at all — cosmochemistry, solid-state physics, thermodynamics, crystallography, astronomy, among others. The consequences are a fair degree of chaos and not a little wasted effort. For as Anderson himself points out, even when people in different disciplines look over each other's shoulders, which is rare enough, they don't always understand the limitations of what they see, with the result that "one discipline's unproven assumptions or dogmas are treated as firm boundary conditions for a theoretician in a slightly overlapping area" (p.xi).

There may be no solution to all this. No one is superhuman, either in the sense of having the ability to comprehend all the relevant subdisciplines or in the sense of having the time and dedication to assimilate completely the bodies of knowledge they represent. On the other hand, much to the envy of the rest of us, there are a few people within the Earth-science community who are, well, fairly superhuman. Don Anderson is one of them — as close to being the compleat geophysicist/geochemist as anyone is ever likely to be.

Theory of the Earth, then, is an extensive summary of practically everything 'known' about the physics, chemistry and physico-chemical evolution of the Earth's interior — the formation of the planet as part of the Solar System; the Earth's bulk composition and the origin of its layering; the physical and chemical characteristics of the crust, mantle and core; the thermodynamics, equations of state, elasticity and solid-state physics of the Earth's interior; the composition and evolution of the mantle and the nature of the geochemical reservoirs therein; the Earth's shape and heat flow; convection in, and

the heterogeneity of, the mantle; mantle phase changes; and the Earth's anelasticity and anisotropy.

These are, so to speak, the broad subheadings, although in themselves they give only a poor idea of what Anderson is about, which is integration. The point here is that the data from any one sub-discipline are often consistent with a whole range of hypotheses but that the possibilities can be narrowed considerably by bringing to bear data from other fields. What we truly know of the Earth's interior could probably be written on the back of an envelope (hence the inverted commas around 'known' in the previous paragraph), but we can reduce the options for uncertainty. This is Anderson's great achievement — to show how evidence from diverse sources can be used to

constrain our ignorance, to rule out the impossibilities and thus reduce the range of possibilities and perhaps even point to probabilities.

This is not really a book for beginners. A great deal of prior knowledge is required (from definitions, through the nature of particular minerals and the use of partial differential equations, to the basic philosophy of Earth-science investigation). There is little about experimental techniques, and in a few places there might just be alternative views. Within these constraints, however, Anderson has produced a remarkable synthesis of our present understanding of the Earth's interior. □

Peter J. Smith is in the Department of Earth Sciences, Open University, Walton Hall, Milton Keynes MK7 6AA, UK.

In the family

P.V.E. McClintock

Near Zero: New Frontiers of Physics. Edited by J.D. Fairbank, B.S. Deaver, Jr, C.W.F. Everitt and P.F. Michelson. W.H. Freeman: 1989. Pp. 959. \$59.95, £47.95.

FEW twentieth-century physicists contribute usefully to more than one fairly specialized area of the subject, and those that do are usually theorists. Landau and Feynman are particularly striking examples but others, too, have found that their theoretical tools developed to deal with one particular problem (say superfluid helium) can also be applied fruitfully in other apparently unrelated areas of the discipline (the structure of neutron stars, for example). The situation for experimental physicists is quite different, in that their skills tend to be less portable. Most of them spend their entire careers working within, or close to, the same narrow specialism that they entered in their twenties. There are some notable exceptions, of course, and none more notable than W.M. (Bill) Fairbank of Stanford University in whose honour *Near Zero* has been compiled.

To many scientists, Fairbank is just "the man with the niobium balls", who reported the seemingly outrageous experimental observations of fractional charge (quarks?) on tiny niobium spheres levitated in a capacitor above superconducting coils. Low-temperature physicists, and many cosmologists and particle physicists, will of course be well aware that his contributions to science have been vastly greater than this one famous (or infamous, depending on your standpoint) experiment. But even they are likely to be surprised by the astonishing scope and breadth of the work

described in this substantial tome.

Compiled by more than 80 contributors, mostly his ex-students and colleagues, the book amounts to a form of scientific biography. It charts in anecdotal detail Fairbank's early life and education; his marriage; his conversion from chemist to physicist; his entry to the world of low-temperature research at Yale; the epoch-making experiments on liquid helium at Duke; and the subsequent extraordinary efflorescence of imagination, creativity and innovation at Stanford, possibly as a reaction to remarks by a blunt friend: "I hear you've got a chance to go to Stanford. You are 42 years old, and I know what happens to such people; I've made a study of it. You get an ultimate position that's very nice at a very good university and you don't do anything more for the rest of your life".

Whether these forthright comments affected the outcome is unclear, but it is certain that no prediction could have been wider of the mark.

Much of Fairbank's work has been guided by the theme of "near zero", which can be construed in various ways. Originally, he used the phrase to refer to experiments near the absolute zero of temperature. Later, it took on a wider meaning as it became clear that frontiers of physics could also be reached by reducing other variables almost to zero, and his experiments have, for example, treated magnetic field, electric field and gravity in just this way. The book has its genesis in a conference of the same name honouring Fairbank's 65th birthday in 1982, for which it effectively constitutes the (somewhat belated) proceedings.

Despite its diverse range of seemingly unrelated topics, and the huge number of contributors, the work as a whole is remarkably unified and coherent. Partly, no doubt, this is the result of careful editing, but it also arises from the fact that the entire corpus is based on the scientific life

and interests of a single individual. The opening sections include reminiscences by the late Walter Gordy on his work with Fairbank both at MIT in wartime and later at Duke. There is also a lengthy, affectionate, wide-ranging and thought-provoking article by C.W.F. Everitt on the nature of the creative process, analysing the origins of Fairbank's imaginative contributions to experimental physics in the context of illustrious predecessors such as Boys, Eötvös and Rutherford.

The bulk of the book, however, is taken up with individual papers on key scientific topics with which Fairbank has been concerned. They are mostly written with an historical/anecdotal flavour by his collaborators of the time in question, and very much from individual points of view, but also bring the subject more up to date and set it in the context of subsequent work. The papers are grouped in chapters covering liquid and solid helium, superconductivity and its applications, fundamental physics (including fractional charge), gravitation and astrophysics, and the surface-shielding effect. Finally, there is a characteristically thoughtful and intriguing chapter by Fairbank himself, presenting his ideas on the future frontiers of physics.

This is very much a family book for a family occasion — not only are Fairbank's close colleagues and collaborators among the contributors, but so too are his brother, the distinguished low-temperature physicist Henry Fairbank, and his laser-physicist son William M. Fairbank Jr, and his wife Jane Davenport Fairbank is one of the co-editors. But is there anything in the book of interest to people outside this immediate family?

The answer must emphatically be yes. All low-temperature physicists will have come across Bill Fairbank or his work in the context of his many contributions to the understanding of liquid helium and superconductivity. Most physicists, and many non-physicists, have heard of the (still highly controversial) fractional charge experiments, of the Stanford Relativity Gyroscope experiment, of the positron free-fall experiments and of the experiments on gravitational waves. All of these topics and many others are carefully described and discussed, mostly at a non-specialist level readily accessible to any physicist. The anecdotal remarks are usually illuminating, often amusing, and only occasionally do they tip over the top in a way that may irritate people outside the family. Anybody interested in low temperatures, or in the nature of the creative process in experimental science, or in the future of physics, should dip into *Near Zero*. They may find themselves reading more of it than they expect. □

P.V.E. McClintock is in the Department of Physics, University of Lancaster, Lancaster LA1 4YB, UK.

Biodigestion

John Edwards

A Guide to Modern Biology: Genetics, Cells & Systems. By Eleanor Lawrence. Longman: 1989. Pp.507. Pbk £9.95, \$29.95.

THE "Modern Biology" of this volume, only partly revealed by its subtitle, is the biology of animal cells, especially where illuminated by recombinant DNA techniques. "Systems" are multicellular: immune, nervous, the developing animal and cancer. Together they claim almost half of the text. Eleanor Lawrence thus sets out single-handedly to tackle territory similar to that covered by six authors in Alberts *et al.*'s *Molecular Biology of the Cell*. She aims, however, to provide a quick reference source, and general introductory reading for the student or non-specialist.

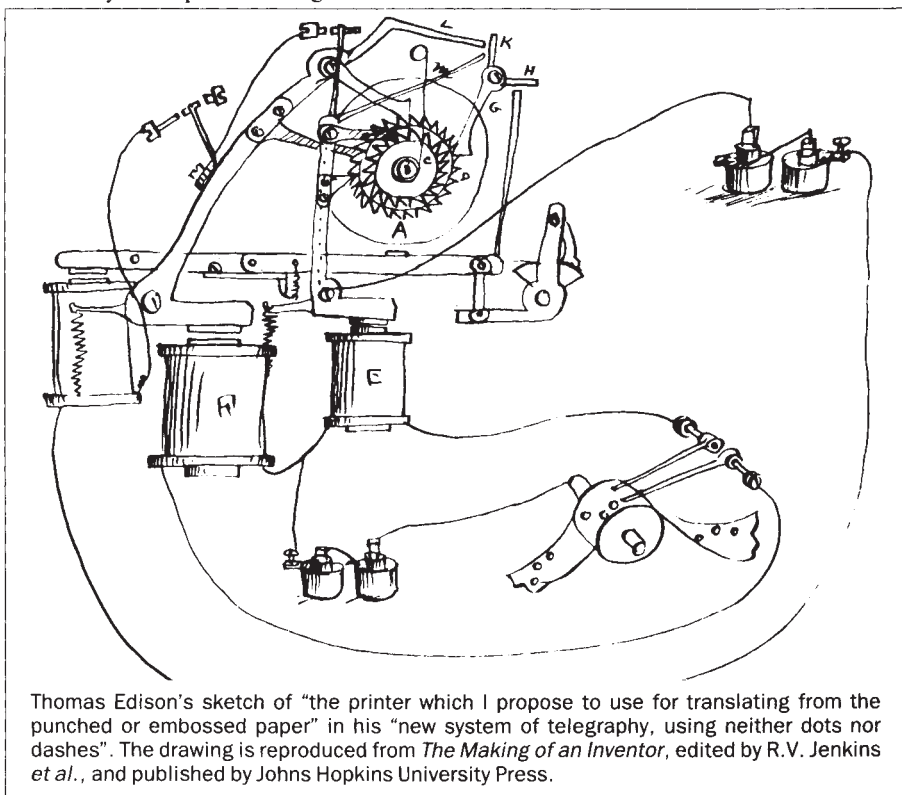
Given the scale of the undertaking, I find that she has been remarkably successful — the key issues are strongly to the fore, essential terms are explained and some of the latest developments are deftly indicated. The crisp English makes for excellent readability. I particularly liked the perspectives in the introductory sections. The sparingly used, clear line diagrams have the virtue of simplicity, although a few carry jargon such as "CAP" and "BFU-E" which remains unexplained. Different views of the IgG molecule are poorly co-ordinated, and a ternary complex linking hormone

receptor to adenylate cyclase gainsays its own legend. Both general and specific references will aid fast entry to the current literature, although they will date rapidly.

No one author could possibly be totally reliable and up-to-the-minute over such a broad canvas. I was unhappy with many odds-and-ends, both biochemical and cell biological (readers with other specialisms would presumably each find a different set of quibbles). To illustrate: there is a curious explanation of why self-splicing in *Tetrahymena* is not properly enzymic; a suspected misinterpretation of GAP with mutant Ras proteins; uncritical reference to the steric model of thin-filament regulation; KDEL but no RGD; and the assertions that all carcinogens are mutagens, and that tumour promoters act by stimulating proliferation of initiated cells. Writing perhaps finished before p53 changed sides from oncogene to anti-oncogene, and focal adhesions from movement to arrest.

We are all non-specialists in relation to some of this material. I believe many biologists will find this volume useful in the way the author intended. Undergraduates may value something easy to carry around, which could also teach them good writing. But for those to whom molecular cell biology is mostly new, the highly resourced artwork, illustrations and wider expertise to be found in a volume such as Alberts *et al.* are not to be missed. □

John Edwards is in the Department of Cell Biology, University of Glasgow, Glasgow G12 8QQ, UK.



Thomas Edison's sketch of "the printer which I propose to use for translating from the punched or embossed paper" in his "new system of telegraphy, using neither dots nor dashes". The drawing is reproduced from *The Making of an Inventor*, edited by R.V. Jenkins *et al.*, and published by Johns Hopkins University Press.

Made to order

John Mann

The Logic of Chemical Synthesis. By E. J. Corey and Xue-Min Cheng. Wiley: 1989. Pp. 436. \$29.95, £19.15.

THE structure elucidation of complex, biologically derived molecules has attracted the attention of organic chemists since the beginning of the nineteenth century. In contrast, the construction of these molecules has been solely a twentieth-century pursuit. Initially, the targets of these endeavours were relatively simple structures; even so, the syntheses were daunting given the paucity of starting materials and chemical reagents, and the complete absence of chromatographic and spectroscopic techniques.

Nevertheless, such complex molecules as quinine and haemin were synthesized before the Second World War. With the advent of increasingly sophisticated spectroscopic and chromatographic methods, and a new understanding of stereochemical principles and of electronic theories of mechanism, ever more challenging targets succumbed to synthesis. By 1960 the growing catalogue of achievements included syntheses of morphine, cortisone, strychnine, colchicine, penicillin V and chlorophyll.

All of these synthetic routes involved several steps, and some of them required heroic efforts in order to overcome poor-yielding stages. Although the routes were chosen after careful consideration of the starting materials available, and with a view to employing particular (often novel) reactions, there is little evidence to suggest that a systematic analysis of the target structures was carried out. The syntheses were thus individual triumphs but not necessarily particularly efficient.

All this changed in 1967, following the publication of a seminal paper by E. J. Corey. He exhorted chemists to analyse their problems, and introduced the concept of retrosynthetic analysis. This involves 'transformation' of the target molecule into a synthetic precursor; and these 'transforms' are just the reverse of the potential synthetic operations. Because there are usually several possible precursors, and because these in turn have numerous other progenitors, a tree of potential intermediates is easily generated. Chemists are then required to exclude those transformations for which the synthesis would (in their opinion) be difficult or impossible. In this way, the shortest route from a cheap or readily available starting material can usually be found.

It would be misleading to suggest that it was Corey alone who invented retrosynthetic analysis; but he was the first to enunciate the general principles, and he

went on to provide a comprehensive set of rules and regulations. The culmination of these enquiries into the logic of synthesis was the design of a computer programme, LHASA, which represents an automated version of the analytical process.

The book reviewed here provides the first full account of 20 years of retrosynthetic analysis. It is divided into three parts. In the first 100 pages, the basic concepts and strategies are explained in full with numerous illustrative examples. This section could easily stand alone as a record of Corey's contributions to the logic of chemical synthesis. In the following 260 pages, retrosynthetic analyses (accompanied by actual synthetic schemes) are provided for 60 biologically derived molecules, and for just about all of the eicosanoids (prostanoids, leukotrienes and so on). All of the syntheses described were carried out by the Corey group. The final section includes a comprehensive literature survey of other multistep syntheses accomplished during the past 20 years.

Rocky choices

K.G. Cox

Origins of Igneous Rocks. By Paul C. Hess. Harvard University Press: 1989. Pp. 336. \$65, £51.95.

IGNEOUS rocks originate as melts in the hot interior of the Earth. They have been erupted or intruded in all geological periods in a great variety of environments, from rift valleys to mountain chains and the ocean floor. Their origins and evolution are highly complex and difficult to study, and it is a brave author who attempts to cover them comprehensively.

Origins of Igneous Rocks is, however, a worthy addition to the recently much-enlarged range of textbooks on this topic. Comparison is invited with A. McBirney's *Igneous Petrology* (Freeman, Cooper 1984), A. Hall's *Igneous Petrology* (Longman, 1987), and M. Wilson's *Igneous Petrogenesis* (Unwin Hyman, 1989). All of these books have a good deal of ground in common — we might ask, do we need a new one, and what is special about this one?

Hess's book is probably the most purely geochemical and theoretically petrological of the four. It has a more extensive section on phase diagrams than the others, and gives good theoretical coverage of magmatic differentiation, principles of geochemistry and the structure of silicate melts. The main part of the book follows the traditional pattern of discussing various rock-associations in sequence, and concludes with an excellent section on lunar petrology, which is not covered in competing works. Hess deliberately does

On one level, the book undoubtedly represents a celebration of Corey's contributions to synthetic and retrosynthetic analysis — and why not! Few other synthetic chemists can claim to have devised and overseen the construction of so many interesting molecules. Some of the schemes have appeared in a similar form in other books (for example in A. Mitra's *The Synthesis of Prostaglandins* published by Wiley in 1977), but to have a full account of theory and practice in one place is invaluable.

This is an advanced-level volume and is not really suitable for undergraduate use. The methodology of retrosynthetic analysis is dealt with in a simpler fashion by Stuart Warren in *Organic Synthesis: the Disconnection Approach* (Wiley, 1982). But for anyone seriously involved in the art of organic synthesis, the book is a must — and a bargain. □

John Mann is in the Department of Chemistry, University of Reading, Whiteknights, Reading RG6 2AD, UK.

not set out to discuss the more 'geological' aspects of the subject, such as igneous rock textures (well done by McBirney) or the forms of intrusions, volcanoes and so on (a strong suit in Hall).

Hess's problem was the sheer difficulty of covering a subject in which there is an immense amount of information about a bewildering variety of rocks, and that of relating the theoretical opening chapters to the subsequent discussion of natural rocks. It can be done, but not in 300 pages. Hess solves the first problem, as others often have, by being less than comprehensive in his examples. Intra-plate volcanism is represented almost entirely by Hawaii, continental flood basalts by the Columbia River province, continental rifts by East Africa, and so on. Wilson is much more comprehensive at this stage, but she devotes a much larger fraction of her book to it.

Like most other authors who have written on igneous petrology, Hess discusses and evaluates existing petrogenetic ideas about specific rock groups. He is particularly good at identifying outstanding questions and inviting readers to form their own judgement, but as a consequence at times can seem almost indecisive. Different authors all see things slightly differently, however, and this may be where the principal value of the new work lies. It contains little that is particularly novel, but it is highly competent and well written. Igneous petrology is a complex subject to become familiar with, and to have another view on most topics is of considerable value. □

K.G. Cox is in the Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK.

Thermodynamic cost of computation, algorithmic complexity and the information metric

W. H. Zurek

Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

Algorithmic complexity is a measure of randomness. In contrast to Shannon's entropy it is defined without a recourse to probabilities; for a binary string s it is given by the size, in bits, of the shortest computer program with the output s . I show that algorithmic complexity sets limits on the thermodynamic cost of computations, casts a new light on the limitations of Maxwell's demon and can be used to define distance between binary strings.

REVERSIBLE computation can be accomplished only by using computer memory to keep track of the exact logical path from the input to the output. This conclusion¹⁻³, is based on the observation that thermodynamic irreversibility is inevitable only in presence of logically irreversible operations⁴⁻⁶: If several input states lead to the same output, the loss of information in such a many-to-one mapping makes it impossible reversibly to 'backtrack' the machinery of the computer. To allow reversible operation the computer must retain this additional information (the history of all the logically irreversible steps) at least temporarily, and it must retain at the end of the computation at least enough extra information to assure unambiguous backtracking^{1,2,6,7}. Thus, reversible computation can be achieved only at the expense of filling up computer memory with historical records, aptly named 'garbage'^{2,6}.

But suppose one insists on replacement of the input with the output in the computer memory: How irreversible would the resulting computation be? This question about the 'replacement computation' leads to simple, yet interesting consequences, among them an information-theoretic measure of distance between binary sequences (which may represent physical configurations) or between statistically characterized ensembles. Thermodynamic efficiency of various information processing operations is of interest in physics of computation¹⁻⁹, and replacement computation is relevant for the foundations of thermodynamics¹⁰.

Consider a simply describable universal computer C . Computations start with some program i as the input, and with the computer in a definite initial state. After the computer stops, only the output o should be present on the output tape and the computer should end up in some unique 'halting state.' As no permanent (only temporary) storage of 'garbage' is allowed, there is no possibility of retaining the history. In general, the computation of o from i will be logically and thermodynamically irreversible.

I shall demonstrate that the least increase of entropy $\delta S(i \rightarrow o)$ associated with the replacement depends on the input and output strings and is at least equal to the difference in length of the shortest programs i^* and o^* to compute i and o on C

$$\delta S(i \rightarrow o) \geq |i^*| - |o^*| \quad (1)$$

Vertical lines denote the size of programs in the number of

symbols. When the computer alphabet is binary the difference of entropy is given in bits.

Equation (1), the first main result of this paper, is derived below, where I also show that the optimum thermodynamic efficiency is difficult to attain: the minimal programs needed to guarantee it cannot be found systematically because of the undecidability of the halting problem—a Turing-machine version of Gödel's theorem.

The total amount of information exchanged in an update $s \rightarrow t$ (the minimal number of erased plus the minimal number of added bits needed to transform file s into file t) provides a natural measure of the information distance between two strings. I also show that this information distance satisfies conditions expected of a metric for both algorithmic and statistical measures of information content. Finally, discussion of the physical significance of the results shows how the logical irreversibility of replacement computations limits thermodynamic efficiency of Maxwell's demon. I propose that algorithmic complexity can be used to define the entropy from the point of view of the demon-like 'intelligent observer', in a manner that takes into account both the lack of complete information and the degree of randomness of the available data about the physical system.

Algorithmic information theory

Algorithmic randomness¹¹⁻²² of a string $K(s)$, also known as algorithmic complexity or algorithmic information content, is defined by the size of the minimal program s^* that, when executed on a universal computer, yields output s

$$K(s) = |s^*| \quad (2)$$

Any message can be represented as a binary string (a sequence of 0s and 1s), but some messages can be reproduced from more concise descriptions: a binary string consisting of 10^6 0s could not be printed in this paper but its content can be (indeed, already has been) communicated to the reader. By contrast, a typical random string of 10^6 0s and 1s obtained by flipping a coin cannot be compressed to a similarly concise description. Incompressible strings are called algorithmically random, and strings that can be characterized by messages small compared to their length in bits are called algorithmically simple.

Equation (2) captures the intuitive difference between the information content and the size of a string (a universal computer is stipulated only to make the definition reasonably independent of the addressee of the message). A string that appears random may nevertheless have a concise description. Binary representations of π , or $\sqrt{2}$, are good examples of apparently random, but, in fact, algorithmically simple strings. In general, no algorithm can distinguish random from simple strings, a difficulty related to Gödel's undecidability^{16-18,22}. As I show below, however, $K(s)$ can be estimated from above by a simple algorithm for any finite s . Moreover, most strings cannot be compressed, so the typical algorithmic randomness of a string is, to leading order, given by its length in bits $|s|$. For example, the algorithmic randomness of a natural number n is typically $\log_2 n$. Certain classes of string are algorithmically random or simple; readers are invited to demonstrate that minimal pro-

grams used in the definition of algorithmic complexity are algorithmically random.

As in Shannon's information theory, one can define algorithmic randomness for sets of strings. Joint algorithmic complexity of a pair of strings (s, t) is the size of the smallest program required to print them both on the output tape. It satisfies the 'commuting' property

$$K(s, t) = K(t, s) + O(1) \quad (3)$$

and the inequality

$$K(s, t) \leq K(s) + K(t) + O(1) \quad (4)$$

The error term of $O(1)$ comes from the slight computer dependence of the algorithmic information content¹³⁻¹⁹. I shall henceforth omit $O(1)$ and replace equality and inequality signs with their approximate versions ($= \rightarrow \approx$; $\geq \rightarrow \gtrsim$, for example) to simplify the notation.

Conditional algorithmic information $K(s|t)$ is the size of the smallest program $q_{s|t}^*$ that can compute string s from the string t (or, equivalently, both s and t from t)

$$K(s|t) \equiv |q_{s|t}^*| \quad (5)$$

Conditional information will be central in our discussion. By analogy with the statistical information theory one would expect $K(s|t)$ to obey

$$K(s, t) = |t^*| + |q_{s|t}^*| \approx K(t) + K(s|t) \quad (6)$$

This equation is satisfied, but only approximately¹⁹. Correction terms of order of $O(\log_2 K(t))$ may arise, with their exact form dependent on the adopted definition of $K(s|t)$. However, equation (6) will hold with no more than $O(1)$ errors if, in addition to string t , one is willing to provide its algorithmic randomness $K(t)$ as the supplementary datum

$$K(s, t) = |t^*| + |q_{s|t, K(t)}^*| \approx K(t) + K(s|t, K(t)) \quad (6a)$$

Moreover, conditional algorithmic randomness of s given t^* (rather than simply t)

$$K(s|t^*) \equiv |q_{s|t^*}^*| \quad (5a)$$

also satisfies equation (6) without logarithmic errors²⁰

$$K(s, t) = |t^*| + |q_{s|t^*}^*| \approx K(t) + K(s|t^*) \quad (6b)$$

Rigorous definition of algorithmic randomness imposes certain requirements on the properties of the minimal programs. In particular, it is convenient to require that minimal programs be self-delimiting¹⁹⁻²¹—they must contain no special symbol to mark their ends, and must therefore contain information about their size. For instance, algorithmic randomness of a typical string defined through a self-delimiting minimal program is given by $|s| + \log_2 |s|$, where the logarithmic term enters because of the necessity to communicate, in addition to the content of s , its size.

Reversibility of compression

What is the least thermodynamic expense of generating string s from its minimal program s^* ? This question emerges in the context of compressing information: by definition, the minimal program s^* can be used to reproduce the more 'verbose' record s , so the operation of replacing s by s^* will preserve the content, but will free $|s| - |s^*|$ bits of memory.

I shall demonstrate that, given s^* , the replacement of s by s^* can be achieved reversibly, with no cost in terms of entropy increase. Note that a reversible program can compute s from s^* and in the process generate history $g(s^*, s)$, which allows it to backtrack from s to s^* reversibly. The process of getting rid of s and replacing it with s^* is:

- Initially, the computer memory contains s and s^* ; The complete input consists of $\{s, s^*\}$.
- The reversible execution of s^* yields output s and history $g(s^*, s)$. The computer memory now contains two copies of s

and $g(s^*, s)$, but does not contain s^* , which was used up, that is, transformed into s in a reversible computation.

■ The newly generated string s can reversibly cancel the 'old' s : $(s, s \rightarrow 0 \dots 0, s)$. This frees up $|s|$ bits, but to make this operation logically reversible one copy of string s must be retained². The memory now contains string s and 'garbage' $g(s^*, s)$.

■ Finally, the computation of s from s^* can be reversed. The computer has all the directions, $g(s^*, s)$, to undo the original computation of s from s^* reversibly.

■ At the end of the operation the computer memory contains only s^* . With no loss of information content, and at no thermodynamic cost, $|s| - |s^*|$ memory bits were liberated.

The string substituting s did not have to be minimal, or even shorter than s : all that was needed was the assumption that the execution of the input s' results in the output s .^{1,2,7}

Lemma 1—Program s' , which generates string s as an output, can be used to reversibly erase s from the memory of the computer, $(s', s) \rightarrow s'$. Minimal program s^* can be employed in this manner to reversibly compress information contained in s .

The essential criterion that makes compression of s possible is that the information in s must be available in the form of some other string s' . So far, I have assumed that the concise s' was already available. To accomplish compression, this s' was used in an algorithm outlined in the proof of Lemma 1. But it is more natural to consider a computer that 'looks for' some convenient and concise s' given the original file s . Lemma 1 then sets the limit on the compressibility of s .

A simple instance of such compressibility occurs when there are two copies of s (ref. 2). Availability of a program s' generating s as an output can be regarded as equivalent to the availability of an implicit copy of s . One can also imagine a more complicated situation where the information needed to generate s resides in a collection of several strings $s', s'', \dots$ which jointly generate s .

The ability of one binary string s to share information content with another string t can be expressed in terms of mutual algorithmic information defined by^{14,20}

$$K(s:t) = K(s) + K(t) - K(s, t) \quad (7)$$

Two strings are algorithmically independent when $K(s:t) \approx 0$. Two copies of the same string are algorithmically redundant: $K(s:s) \approx K(s)$. Moreover, string s is algorithmically redundant with respect to string t when

$$K(t) > K(s) \quad (8a)$$

and

$$K(s:t) \approx K(s) \quad (8b)$$

or equivalently

$$K(s, t) \approx K(t) \quad (8c)$$

In other words, s is algorithmically redundant with respect to t when it contains no additional information not already included in t . In this sense, the output must be redundant with respect to the input.

Reversibility of the replacement operation $(s', s) \rightarrow s'$ relies on redundancy of the output string s with respect to the input s' . Lemma 1 is an example of the more general fact established in ref. 1: if $s \rightarrow f(s)$ is an arbitrary computable function, operations $s \rightarrow (s, f(s))$ and $(s, f(s)) \rightarrow s$ can be carried out reversibly because the inputs and the outputs are mutually redundant. What is new here is the application: Lemma 1 uses algorithmic complexity to establish the limit on the thermodynamically reversible compressibility of information.

Is redundancy not just a necessary but also a sufficient condition for the reversibility of the replacement operation $(s', s) \rightarrow s'$? It is true that whenever $K(s', s) - K(s) \approx 0$ there must exist, by definition of conditional algorithmic information, a short minimal conditional string $q_{s|s'}^*$ capable of turning s' into a

program with output s . But the existence of a concise program generating s does not guarantee its availability because minimal programs are not recursively computable functions of their respective outputs^{13–20}. The possibility of a reversible compression of s into the s^* or the existence of $q_{s|s}^*$, thus establishes limits on thermodynamic efficiency, but does not guarantee that optimal efficiency can be attained. Nevertheless, the absolute thermodynamic cost of computations is of interest for the foundations of statistical mechanics and, as demonstrated in more detail below, can be settled through arguments involving existence rather than computability.

Thermodynamic cost of replacement computation

I now investigate the least thermodynamic price (the optimal attainable efficiency) of a general replacement computation $i \rightarrow o$. The output is redundant with respect to the input, and therefore typically contains less information. The replacement computation of o from i thus disposes of some information, and must be logically irreversible. These logically irreversible steps can be either 'paid for' at once, the history of the computation being erased as the computation is carried out, or reversibly recorded, which delays thermodynamically costly erasures until after the computation is completed.

The cost of erasures will not increase if they are delayed by making a record $g(i, o)$ of all the irreversible steps and erasing it after the computation is completed—a bit of information is erased at the same cost regardless of when the erasure is done. One can therefore consider reversible computations $i \rightarrow o, g(i, o)$ and, without loss of generality, establish the lower bound on their efficiency by finding the least thermodynamic cost of erasure of $g(i, o)$ in the presence of o . Because the history will be generally algorithmically non-random and may contain information already present in the output, any concise program $q_{i|o}$ that can reproduce i from o and thereby reconstruct the history $g(i, o)$ can be also used, by Lemma 1, to reversibly compress $g(i, o)$. This line of reasoning can be developed into:

Theorem 1—The thermodynamic cost of computation which turns input i into output o , retaining no memory of i or of the intermediate steps cannot be made lower than

$$\delta S(i \rightarrow o) = K(i|o) \quad (9)$$

where $K(i|o)$ is the size of the minimal conditional string $q_{i|o}^*$.

The proof is accomplished by contradiction. Let the history $\tilde{g}(i, o)$ allow computation of o from i , and suppose that it contains fewer than $K(i|o)$ bits, $|\tilde{g}(i, o)| < K(i|o)$. Then $\tilde{g}(i, o)$ could be used to reproduce i from o with a program shorter than the minimal conditional program $q_{i|o}^*$, which is impossible. Therefore $q_{i|o}^*$, containing $K(i|o)$ bits, is the shortest possible record of the history.

Erasure of the conditional string cannot be accomplished at a cost of less than $\delta S(i \rightarrow o) = |q_{i|o}^*|$ bits of entropy, which demonstrates Theorem 1. It also establishes:

Corollary—There exists an abbreviated version $q_{i|o}^*$ of the history of the computation $i \rightarrow o$ with $K(i|o)$ bits. It is typically much shorter than the original history $g(i, o)$ but, like g , it enables i to be computed from o and can hence be used to reversibly reconstruct g itself. Unlike g , however, it is not necessarily computable from the input.

Uncomputability means that although Theorem 1 and its corollary establish a firm lower bound on the size of the history and demonstrate that this bound can be (not by recursive computation, but perhaps by sheer luck) actually met, they offer little help in the task of minimizing thermodynamic expenditures.

The lower bound of Theorem 1 can be used to derive equation (1), a slightly weaker but more direct bound expressed in terms of the algorithmic complexities of i and o . Because the output is redundant with respect to the input, $K(i, o) \approx K(i)$, and from equation (6a), $K(i|o, K(o)) \approx K(i) - K(o)$. The size of the minimal program that computes i from o alone cannot be less than $K(i) - K(o)$; $K(i|o) \geq K(i|o, K(o))$, so:

Lemma 2—The thermodynamic cost of a replacement computation is greater than the difference in the algorithmic information content between input and output

$$\delta S(i \rightarrow o) \geq K(i) - K(o) \quad (10)$$

This simple equation gives the lower bound in terms of the individual complexities of the two strings. Moreover, it is the absolute limiting efficiency for the case when o^* , the minimal program for the output, is at hand. This follows from equation (6b) for the conditional information given o^* : $K(i|o^*) \approx K(i) - K(o)$ without logarithmic corrections.

The disadvantage of equation (10) is that it is in general somewhat weaker than Theorem 1, which contained directly the size of the minimal conditional string. The reason for this slight difference touches some of the fascinating (and somewhat paradoxical) aspects of algorithmic information theory. Consider computation of a string s from its minimal program s^* . The algorithmic randomness of s and s^* is almost ($\sim O(1)$) identical: $K(s) \approx |s^*|$. Moreover, s^* is algorithmically random, as the reader verified above. Hence, the shortest program needed to obtain s^* cannot be shorter than $K(s)$. In addition, it cannot be longer than $K(s)$ by more than a few bits, as it must be shorter than the binary version of 'Print s^* '. Consequently, $K(s) \approx K(s^*)$. Yet, s^* cannot be computed from s in a finite number of steps because of gödelian undecidability^{16–18,22}.

This comment should not be misunderstood. It is not difficult to suggest a concise and straightforward method of finding the shortest program that will generate string s within a pre-determined (arbitrarily large, but finite) number of steps N . One can test on C all of the conceivable programs (all the strings) shorter than the binary version of some program that obviously generates s . ($s_p \equiv$ Print s with the self-delimiting correction, if necessary, would do the job.) The list

0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, 101, 110, 111, . . .

of all binary strings shorter than $|s_p|$, which must include all candidate programs, is very large ($\sim 2^{|s_p|}$) but finite, and can be easily supplied by a concise algorithm.

Each of the candidate programs can be run for N steps in a finite time $\leq N \times 2^{|s_p|}$. Among the programs that have halted within N steps some may give an output s . The first of them in the lexicographic order, $\tilde{s}$, is easily identified, and if $\tilde{s} < s$ it can be used to accomplish a partial compression of the information. If none of the programs that have halted produces s , s_p is still the shortest known program to generate s , and information compression has not been found possible.

This algorithm can be implemented reversibly, so the thermodynamic cost of finding $\tilde{s}$ is, in principle, negligible. But this does not disprove the fundamental undecidability of the algorithmic randomness of binary strings^{16–18,22} because the algorithm used to find $\tilde{s}$ is primitive recursive^{23,24}; it must halt after a finite number of steps $\leq N \times 2^{|s_p|}$. Identity of $\tilde{s}$ with s^* could be established only in the $N \rightarrow \infty$ limit.

Indeed, if it were known beforehand which programs never halt, they could be eliminated from the lexicographic list. Now the shortest remaining program with output s would be guaranteed to yield the correct value of $K(s)$, and the minimal program s^* could be easily found. This clever plan cannot be used: Turing's halting theorem²⁵, the computer-theoretic equivalent of the Gödel's theorem, establishes that there is no way to distinguish halting from non-halting programs. Undecidability thus thwarts attempts to systematically derive, in a finite number of steps, the most efficient algorithm. Worse than that, the minimal thermodynamic cost of computation is given in terms of algorithmic complexities, which are themselves uncomputable. Not only is it impossible to find the best procedure for optimizing a computation, it is also impossible to find out how well one is doing.

Advance knowledge of the algorithmic complexity $K(s)$ of s would allow one to modify the search so that it would be

guaranteed to halt after a finite number of steps. A short program for s , equal in length to the minimal s^* , would therefore become recursively computable. The uncomputability of the optimal efficiency is thus essential for the consistency of the above arguments, but it remains true that infinite computations can discover minimal programs.

The optimal algorithm could conceivably be found quickly. The halting theorem does not disallow the best solution being found after a finite number of iterations. One might also find the best solution by a lucky guess. It is only certain that one will never know how lucky one was.

Although the optimally efficient algorithm cannot be systematically discovered, it is always possible to limit the number of the erased bits to the size of the input^{1,2,7}

$$\delta S(i \rightarrow o) \leq |i| \quad (11)$$

This 'guaranteed efficiency' can be easily justified¹: it is always possible to perform the computation $i \rightarrow o$, $g(i, o)$, duplicate o , and use $g(i, o)$ to reproduce i . One is now left with o and i . Erasure of i can always be carried out at the cost of $|i|$ bits, so the thermodynamic cost can always be limited to no more than $|i|$ bits.

Although the δS guaranteed by equation (11) is somewhat trivial, there are instances when it coincides with the lower bound given by Theorem 1 and Lemma 2. For example, replacement of a random string with an equally long string of zeros cannot be carried out at a price lower than the guaranteed efficiency, equation (11).

These results complement and extend the proof¹ that a computation can be always performed reversibly^{1-3,5-9} by establishing the minimal price of such reversibility for replacement computations. The price is given by the algorithmic complexity of the history of the computation given its output. Retaining $g(i, o)$ or i , leaves one the option to undo the computation at no thermodynamic expense. Erasing the history and the input incurs at least $\delta S(i \rightarrow o) = |q_{i|o}^*| \geq K(i) - K(o) = |i^*| - |o^*|$ of entropy increase.

One might argue that the distinction between these two options—erasure and storage—is artificial: erasure can be accomplished by storing the relevant information in a physical system outside the computer memory. This point of view, which identifies increase of entropy with the erased information, must be further extended in a discussion of an update. Consider two files, s and t , each containing some information not present in the other $K(s) \neq K(s:t) \neq K(t)$. To compute t from s one must supply conditional information $K(t|s)$. A replacement computation requires erasure of the spurious information $K(s|t)$.

It is natural to count the thermodynamic cost of the supplied information with the opposite sign from the cost of the erased information. This leads to the net flow of algorithmic information between the computer and the rest of the Universe

$$\delta K = K(s|t) - K(t|s) \quad (12)$$

which (with the logarithmic correction terms) could be reduced to the right hand side of equation (10). But δK cannot immediately be identified with the minimal dissipation. The arguments of ref. 4 imply only that erasure of information is associated with an increase of entropy. δK would be a valid expression for the thermodynamic cost of an update only if the transfer of information from outside the computer caused a corresponding decrease of entropy. This assumption may be not unreasonable, (acquisition of information can be regarded as a measurement, and measurements are thought to decrease statistical entropy) I shall not attempt to discuss conditions for its validity here. Moreover, neither of the two ingredients of δK is recursively computable, so the actual net number of transferred bits could be larger or smaller than δK . Thus, these arguments do not allow δK , equation (12), to be regarded as a minimal thermodynamic cost in the sense of equation (1) and Theorem 1.

I shall return to the thermodynamic significance of replacement computation later, in a discussion of measurements and the laws of thermodynamics. A more complete analysis of related issues can be found in ref. 10.

Algorithmic distance and the information metric

The extent of the update can be measured by the least amount of information involved in transforming s into t . Instead of counting the $K(s|t)$ and $K(t|s)$ with signs that depend on the direction of their flow between C and the rest of the Universe, as was done to obtain equation (12), one can add them to establish the total amount information needed to carry out the replacement. This sum defines the information distance between s and t

$$\Delta(s, t) \equiv K(s|t) + K(t|s) \quad (13a)$$

Algorithmic information distance can also be approximately expressed as

$$\Delta'(s, t) \equiv K(s) + K(t) - 2K(s:t) \quad (13b)$$

$$\equiv K(s, t) - K(s:t) \quad (13c)$$

$$\equiv 2K(s, t) - K(s) - K(t) \quad (13d)$$

The equality between $\Delta(s, t)$ and $\Delta'(s, t)$ would be exact only if equation (6) were satisfied exactly. For statistical (Shannon) entropy, analogues of Δ and Δ' are identically equal, but logarithmic correction terms in the algorithmic version of equation (6) imply similar (logarithmic) corrections going from the original definition, equation (13a), to its alternatives, equations (13b-d). One could get rid of these corrections by using one of the alternative but less natural definitions of conditional algorithmic information (equation (6a) or (6b)).

The information-theoretic quantity $\Delta(s, t)$ defined by equation (13a) satisfies the three conditions required to allow its use as a distance; (i) positivity

$$\Delta(s, t) \geq 0 \quad (14)$$

with the equality holding only when $s = t$, (ii) symmetry

$$\Delta(s, t) = \Delta(t, s) \quad (15)$$

and (iii) the triangle inequality

$$\Delta(s, t) + \Delta(t, u) \geq \Delta(s, u) \quad (16)$$

which follows from noting that the conditional minimal program $q_{u|s}^*$ for computing u from s cannot be longer than the program that computes u from s using t (that is, the concatenated program $q_{u|t}^* q_{t|s}^*$). Consequently

$$K(u|t) + K(t|s) \geq K(u|s) \quad (17a)$$

The same argument holds for computation of s from u . Thus

$$K(t|u) + K(s|t) \geq K(s|u) \quad (17b)$$

Addition of these two inequalities with the help of equation (13a) yields the triangle inequality, equation (16). Hence:

Theorem 2—Information distance $\Delta(s, t)$ provides a metric for the space of binary sequences.

Another proof of the triangle inequality, more immediately applicable to the statistical (Shannon) version of the information distance, follows from

$$K(s, t, u) \geq K(s, u) \quad (18)$$

Both sides of equation (18) can be multiplied by two, and after adding $K(s, t) - K(s, t) + K(t, u) - K(t, u) = 0$ to the left-hand side and subtracting $(K(s) + K(u))$ from both sides one arrives at $K(s|t, u) + K(t|s) + K(t|u) + K(u|s, t) \geq K(s|u) + K(u|s)$. Now, ignoring the usual logarithmic correction terms and using the fact that the left-hand side of the above inequality cannot decrease if $K(s|t, u)$ and $K(u|s, t)$ are exchanged with $K(s|t)$ and $K(u|t)$, respectively, leads to the triangle inequality in the

form in which Δ is expressed through conditional information, equation (13a).

All the steps of this second proof apply, without any logarithmic corrections, to Shannon's entropy H of classical ensembles, S , T and U . There is then no difference between Δ and Δ' , and the distance given by any of the versions of equation (13) provides a metric for a collection of statistical ensembles characterized by their probability distributions. But it should be noted that the 'self-evident' inequality, equation (18), should not be taken for granted: it is not satisfied for statistical entropy H for systems that exhibit non-separable quantum correlations.

It is not difficult to show that $\Delta(s, t)$ satisfies the inequality $K(s, t) \geq \Delta(s, t) \geq |K(s) - K(t)|$. Obvious applications of the information distance include characterization of the relationship of the probability distributions. To some degree, mutual information has been already employed in that role²⁶. Algorithmic-information distance is perhaps less immediately applicable to practical issues, but its fundamental implications may be far-reaching.

Complexity, Maxwell's demon and entropy

We have found that replacement computation can be done at a thermodynamic expense (an increase of entropy) greater than or equal to the difference in the algorithmic randomness between input and output. Entropy increase is thus a direct consequence of the decrease in the algorithmic information content. So far, this relation has been applied to set the thermodynamic limit on the reversibility of replacement computation. But there is also an inverse problem: what limits on statistical mechanics and thermodynamics can be found from algorithmic concepts? This is discussed in detail elsewhere¹⁰; here, I shall briefly outline the basic argument to provide physical motivation for interest in algorithmic randomness.

Consider a computerized engine consisting of a physical system in contact with a thermal reservoir at temperature T and controlled by universal computer C . The system is initially described by some density matrix ρ . Its standard ensemble entropy, measured in bits, is given by

$$H(\rho) = -\text{Tr } \rho \log_2 \rho \quad (19)$$

The computer initiates a cycle of the engine by performing a dissipationless measurement^{2,10,27} on the system. (In the context of quantum theory this implies that the measured observable must commute with the entropy, a condition which is always satisfied in the classical context.) Each of the outcomes, which are assumed to be mutually exclusive, is described by a density matrix ρ_k and occurs with probability p_k , so that $\rho = \sum_k p_k \rho_k$. The entropy of each outcome $H(\rho_k)$ is smaller than $H(\rho)$, and the average decrease of entropy is equal to

$$\langle \Delta H \rangle = H(\rho) - \sum_k p_k H(\rho_k) = -\sum_k p_k \log_2 p_k \quad (20)$$

The computer can convert this gain of information into useful work, at no thermodynamic expense, by constraining the system to the part of the phase space that coincides with ρ_k . Subsequently, useful work

$$\Delta W^+ = T(H(\rho) - H(\rho_k)) = T\Delta H_k \quad (21a)$$

could be extracted by allowing the constraint to relax in a slow isothermal fashion, so that ρ_k is transformed into ρ (Boltzmann's constant $k_B = 1$). The cycle can be repeated. If the process were truly cyclic, it would result in a violation of the second law by allowing an average gain $T\langle \Delta H \rangle$ of useful work per cycle. Szilard's engine provides a gedanken experiment realization of this idea^{2,27-29}.

Although the computerized "Maxwell's demon" can seemingly extract useful work, the second law of thermodynamics is not in danger. The cycle is never completed^{2,10,27-29}, as the memory of the computer does not return to its initial state:

useless information about the outcomes of the past measurements piles up. The process uses computer memory as a zero-entropy reservoir. To make the process truly cyclic, the memory has to be periodically erased, and the cost of erasure must be subtracted to calculate the actual amount of useful work extracted. When the data is not random, a 'smart' erasing mechanism can take advantage of the regularities and compress the information before the erasure, whereas a 'dumb' erasing mechanism incapable of information compression would have to use $k_B T$ of work per bit, regularities notwithstanding. Data compression can be done separately, however, before the result is passed on to the 'dumb' erasing mechanism. One can thus separate reversible compression from the final irreversible erasure, and the cost of erasure can be in principle made as low as $k_B T$ times the algorithmic information content, rather than size in bits, of the record in question. This efficiency is attained by prior maximum (reversible) compression of the data to be erased (that is, in case of the computerized demon, of the conditional information $K(\rho_k|\rho)$). From Theorem 1 and Lemma 2, the thermodynamic cost of the most efficient erasure cannot be less than

$$\Delta W^- \approx T\Delta K_k, \quad (21b)$$

where $\Delta K_k = [K(\rho) - K(\rho_k)] < 0$ is the difference between the algorithmic information content of the initial state and of the measurement outcome k .

Here I have used the version of the lower limit, $\delta S(\rho_k \rightarrow \rho) = K(\rho_k) - K(\rho)$, appropriate for a computation starting with the description of ρ_k as the input and ending with the minimal description of ρ as the output. This shows that even if the demon was able to guess minimal programs, it would still not endanger the second law. Without minimal descriptions, the cost of erasure would increase and the efficiency would decrease. The second law is therefore satisfied for computer-operated engines independently of the issue of computability, but uncomputability of minimal strings will cause additional dissipation.

The increase in the potential to do useful work as a result of the information gain¹⁰ must be measured not just by the decrease in the ensemble entropy ΔH_k , but by the sum of ΔH_k and ΔK_k , which gives the least number of bits needed to write down the outcome of the measurement. The net amount of work that can be gained following measurement with the outcome k is

$$\Delta W_k = \Delta W_k^+ + \Delta W_k^- = T(\Delta H_k + \Delta K_k) \quad (22)$$

Consequently, what determines the net amount of extractable work is the sum of the usual ensemble entropy, which specifies the missing information, plus the algorithmic randomness of the available information¹⁰

$$\mathcal{S} = H + K \quad (23)$$

Physical entropy $\mathcal{S}$ extends the applicability of arguments based on entropy beyond the usual ensemble setting by allowing discussion of definite outcomes of measurements and their consequences¹⁰. In the usual limit it is dominated by the ensemble Boltzmann-Gibbs-Shannon component H . When the state of the system is known sufficiently well, however, algorithmic randomness K supplies the key contribution. Moreover, only when this state is algorithmically simple can one extract a sizable amount of useful energy.

The second law formulated in terms of physical entropy, equation (23), holds even during the measurements because the average information gain $\langle \Delta H \rangle$ is offset in $\mathcal{S}$ by the least average size of the code $\langle \Delta K \rangle = \sum_k p_k \Delta K_k$ required to specify the outcome. Information and coding theory implies that^{30,31}

$$\langle \Delta H \rangle \leq \langle \Delta K \rangle \quad (24)$$

So, recording the outcome of the measurement performed on a system in a thermodynamic equilibrium will, on the average, use up at least as many memory bits as the corresponding decrease of ignorance ΔH . Fluctuations with $|\Delta K_k| < \Delta H_k$ are possible, but they occur sufficiently infrequently that equation

(24), and the second law formulated in terms of $\mathcal{S}$ are satisfied for the ensemble. Erasure of information can be represented, in terms of the $\mathcal{S}$, as the act of storing the information from the memory of the computer in some external system. This point of view was mentioned earlier.

Theorem 1 establishes an interesting connection between algorithmic complexity and the thermodynamic depth recently proposed as a measure of complexity³². Thermodynamic depth of an object o is the amount of information discarded in course of the process that led to its formation. As defined, thermodynamic depth is a function of a specific history³², but it is of interest to define a minimal thermodynamic depth that is solely a function of the initial conditions i and of the final result o . Theorem 1 demonstrates that such minimal thermodynamic depth is given by the size of the smallest program that can reconstruct i from o and is closely approximated by the difference of their algorithmic information contents. Further discussion of the applications of algorithmic concepts to complexity can be found in ref. 33.

Conclusions

The central idea in this paper is a computational counterpart of the second law of thermodynamics. A computation that replaces the input (program) file with the output causes an increase of entropy by no less than $\delta S(i \rightarrow o) \approx |q_{i|o}^*| \geq K(i) - K(o)$ bits, where $q_{i|o}^*$ is the minimal conditional string, the shortest conceivable version of history of the logically irreversible transformation $i \rightarrow o$. This implies that a compression of information, replacement of a file s with a known program $\hat{s}$ that computes s , can be accomplished reversibly, and could save as much as, but no more than $|s| - K(s)$ bits of memory.

The second law of algorithmic dynamics sets limits on the efficiency of Maxwell's demon: Even if the demon could acquire the relevant information about the state of the physical system without dissipation, and correctly guess the most concise method of representing its knowledge, the necessity to periodically erase from its memory the records of 'used up' past measurements implies that at least $\delta S \geq K(\text{measurement outcome}) - K(\text{initial$

state) bits would have to be erased in each successful cycle. This logical irreversibility of replacement computations is enough to guarantee the safety of the second law of thermodynamics even in the presence of information gathering and utilizing systems.

The demon's inability to circumvent the second law by any combination involving measurements and computation is especially obvious when the process of energy extraction is discussed in terms of physical entropy defined as a sum of known disorder (measured by the algorithmic randomness) and remaining ignorance (measured by Shannon's entropy).

Smoluchowski³⁴, in his discussion of the famous 'trapdoor' (an automated version of Maxwell's demon) showed that "no automatic, permanently effective perpetual-motion machine" can violate the second law by taking advantage of statistical fluctuations, but he suggested that "such device might perhaps function if it were operated by intelligent beings".

I have demonstrated that the second law is safe even from 'intelligent beings', as long as their abilities to process information are subject to the same laws as these of universal Turing machines. (This is the so-called 'Church-Turing thesis', which can be succinctly restated as²⁴ "What is human-computable is also machine-computable".) Moreover, it is unlikely that a demon dealing with an equilibrium system will 'break even', because optimal efficiency of replacement computations can be attained only with the necessary minimal programs (s^* , $q_{i|o}^*$, for example). Unfortunately, minimal programs cannot be systematically computed—Turing's halting theorem implies that the information required to attain maximum efficiency can be secured only through an infinitely long computation. Thus, Gödel's undecidability can be regarded as an additional source of dissipation.

Algorithmic distance between binary strings can be defined as a difference between their joint and mutual complexities. It has all the properties required of a metric. Moreover, an analogous formula—the difference between Shannon's joint entropy and mutual information—can be used to define distances between statistical ensembles. $\square$

Received 14 September 1988; accepted 2 August 1989.

1. Bennett, C. H. *IBM J. Res. Dev.* **17**, 525–532 (1973).
2. Bennett, C. H. *Int. J. theor. Phys.* **21**, 305–340 (1982).
3. Bennett, C. H. *IBM J. Res. Dev.* **32**, 16–25 (1988).
4. Landauer, R. *IBM J. Res. Dev.* **3**, 113–131 (1961).
5. Landauer, R. in *Signal Processing* (ed. Haykin, S.) 18–47 (Prentice-Hall, New York, 1989).
6. Bennett, C. H. & Landauer, R. *Scient. Am.* **253**, 48–56 (1985).
7. Fredkin, E. & Toffoli, T. *Int. J. theor. Phys.* **21**, 219–253 (1982).
8. Zurek, W. H. *Phys. Rev. Lett.* **53**, 391–394 (1984).
9. Bennett, C. H. *SIAM J. Comput.* **18**, 766–776 (1989).
10. Zurek, W. H. *Phys. Rev. A* (in the press).
11. Solomonoff, R. J. *Inf. Control* **7**, 1–22 (1964).
12. Kolmogorov, A. N. *Inf. Transmission* **1**, 3–11 (1965).
13. Kolmogorov, A. N. *IEEE Trans. Inf. Theory* **14**, 662–664 (1968).
14. Kolmogorov, A. N. *Usp. mat. Nauk* **25**, 602–613 (1970).
15. Chaitin, G. J. *J. Ass. Comput. Mach.* **13**, 547–569 (1966).
16. Chaitin, G. J. *Scient. Am.* **232**(5), 47–52 (1975).
17. Chaitin, G. J. *IBM J. Res. Dev.* **21**, 350–359 (1977).
18. Chaitin, G. J. *Algorithmic Information Theory* (Cambridge University Press, 1987).
19. Gacs, P. *Soviet Math. Dokl.* **15**, 1477–1478 (1974).
20. Chaitin, G. J. *J. Ass. Comput. Mach.* **22**, 245–268 (1975).
21. Levin, L. A. *Soviet Math. Dokl.* **17**, 522–526 (1976).

22. Stewart, I. *Nature* **332**, 115–116 (1988).
23. Davies, M. *Computability and Undecidability* (McGraw-Hill, 1958).
24. Hofstadter, D. *Gödel, Escher, Bach* (Random House, New York, 1979).
25. Turing, A. M. *Proc. Lond. math. Soc.* **42**, 230–265 (1936).
26. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes* Section 13.6 (Cambridge University Press, 1987).
27. Bennett, C. H. *Scient. Am.* **255**(11), 108–117 (1987).
28. Szilard, L. *Z. Phys.* **53**, 840–856 (1929).
29. Zurek, W. H. in *Frontiers of Nonequilibrium Statistical Mechanics*, (eds More, T. G. & Scully, M. O.) 151–161 (Plenum, New York, 1986).
30. Shannon, C. E. & Weaver, W. *The Mathematical Theory of Communication* (University of Illinois Press, 1949).
31. Hamming, R. W. *Coding and Information Theory* (Prentice-Hall, Englewood Cliffs, 1987).
32. Lloyd, S. & Pagels, H. *Ann. Phys.* **188**, 186–213 (1988).
33. (ed. Zurek, W. H.) *Proc. Santa Fe Inst. Workshop "Complexity, Entropy, and Physics of Information"* (Addison-Wesley, in the press).
34. Smoluchowski, M. in *Vorträge über die Kinetische Theorie der Materie und der Elektrizität* (Teubner, Leipzig, 1914).

ACKNOWLEDGEMENTS. I would like to thank C. H. Bennett for comments and suggestions, M. Gell-Mann and S. Lloyd for discussions, and the Santa Fe Institute and the Institute for Theoretical Physics in Santa Barbara for hospitality. This research was supported in part by the NSF and the National Aeronautics and Space Administration.

Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis

Joachim Ostermann, Arthur L. Horwich*, Walter Neupert & F.-Ulrich Hartl

Institut für Physiologische Chemie der Universität München, Goethestrasse 33, 8000 München 2, FRG

* Yale University School of Medicine, Department of Human Genetics, 333 Cedar Street, New Haven, Connecticut 06510, USA

Mitochondrial heat-shock protein hsp60 functions in the folding of proteins imported into mitochondria. Folding occurs at the surface of hsp60 in an ATP-mediated reaction, followed by release of the bound polypeptides. We propose that hsp60 catalyses protein folding.

NEWLY synthesized cytosolic precursors of mitochondrial proteins have to maintain an unfolded conformation to be competent for membrane translocation^{1,2}. This is probably achieved by interaction with heat-shock proteins of relative molecular mass 70,000 (70K) in the cytosol and by the action of other factors which might include ATP-dependent 'unfolding enzymes'³⁻⁹. Precursors traverse the mitochondrial membranes in an extended conformation at translocation contact sites between the outer and inner membranes¹⁰⁻¹⁴. Amino-terminal presequences are cleaved by a matrix-localized, metal-dependent processing enzyme¹⁵⁻¹⁷. Proteins remaining in the matrix compartment then have to refold, and in many cases to assemble into supramolecular complexes. The precursors of a number of proteins of the intermembrane space are re-exported across the inner membrane, and follow an evolutionarily conserved assembly pathway of procaryotic origin^{18,19}. These proteins probably have to remain in a more loosely folded conformation in the matrix before the second membrane-translocation event.

Little is known about how proteins fold inside cells. The mechanisms underlying the folding and assembly of proteins imported into the mitochondria are now a focus of interest. The recently identified stress protein, hsp60, is the first mitochondrial component known to be essential in these processes^{20,21}, which were previously assumed to occur spontaneously²². Hsp60 is a nuclear-coded, constitutively expressed heat-shock protein residing in the mitochondrial matrix as an oligomer of 14 subunits^{23,24}. The hsp60 equivalent has been detected in mitochondria from several sources, including those of human cell lines²⁵⁻²⁷. Together with the structurally related *Escherichia coli* heat-shock protein groEL and the α -component of the Rubisco subunit-binding protein, hsp60 belongs to a subclass of molecular 'chaperones' termed 'chaperonins', which are components assisting in oligomeric protein assembly by an unknown mechanism^{25,28,29}. The consequences of loss of hsp60 function have been analysed in a temperature-sensitive lethal yeast mutant defective in the gene coding for hsp60²⁰. Mutant cells are deficient in the assembly of several mitochondrial proteins of the matrix, inner membrane and intermembrane space. For example, the precursor of the β -subunit of F_1 -ATPase is imported into the mutant mitochondria, but fails to assemble into the F_0F_1 -ATPase. Proteins of the intermembrane space, such as cytochrome b_2 , which after import into the matrix have to be re-translocated across the inner membrane, do not reach their target compartment.

Using mitochondria as a model 'cell', we investigated whether the folding of imported proteins *per se* and not only the assembly

of subunits into supramolecular complexes could be mediated by proteins such as hsp60. Here we describe details of the molecular folding of proteins transported into the mitochondrial matrix. At low temperatures, after depletion of mitochondrial ATP, or after treatment of mitochondria with *N*-ethylmaleimide, the precursor of a mitochondrial fusion protein is translocated across the mitochondrial membranes but does not assume a folded conformation. Intermediates arrested in the unfolded state are associated with hsp60 in a high-molecular weight complex. At least partial folding occurs in an ATP-dependent reaction at the surface of hsp60 before release. Essentially the same pathway is followed by authentic mitochondrial proteins. We conclude that (re)folding of proteins after translocation across the mitochondrial membranes requires a proteinaceous machinery in the matrix: hsp60 is identified as an essential component of this machinery which appears to fulfill the function of an ATP-dependent 'folding catalyst'.

Folding intermediates

To analyse the pathway of (re)folding of newly imported proteins, we used a mitochondrial precursor protein whose state of folding could be monitored by virtue of its resistance to protease. We made use of the fact that the cytosolic enzyme dihydrofolate reductase (DHFR) retains its enzymatically active conformation and folds as an independent, highly protease-resistant unit when fused to the pre-sequence of a mitochondrial precursor^{13,30}. The fusion protein preSu9-DHFR proved to be a suitable construct: it consists of amino-acid residues 1-69 of the precursor of *Neurospora* F_0 -ATPase subunit 9, joined to the amino-terminus of the complete mouse DHFR by three linker residues³¹. The 66-amino acid presequence is cleaved in two steps by the mitochondrial processing peptidase^{32,33}. Unfolding of the DHFR moiety is the rate-limiting step for the transport of similar fusion proteins across the mitochondrial membranes^{13,34}. To achieve rapid membrane translocation, radiolabelled preSu9-DHFR was precipitated from a reticulocyte lysate with ammonium sulphate, and the precipitate dissolved in 8M urea. After dilution in the import incubation, this precursor is rapidly translocated into mitochondria from *N. crassa* in a membrane-potential-dependent manner. Imported precursor was processed to the mature form, which corresponds to the complete DHFR carrying six additional residues at the amino-terminus. To determine the kinetics of refolding, import reactions were terminated after a short incubation by addition of uncoupler with immediate cooling. Precursor associated with the surface of mitochondria was removed by treatment with proteinase K. The mitochondria were then permeabilized by digitonin and the state of folding of the imported protein assayed by measuring the protease-resistance of its DHFR component. About 90% of the preSu9-DHFR contained in the reaction is translocated into mitochondria within 45 s at 25 °C (Fig. 1a), but only 30% of the imported protein reaches the stably folded, protease-resistant conformation. After three minutes, 70% of the imported protein has folded. Lowering the temperature considerably slows down the refolding and proteolytic processing of imported Su9-DHFR,

although the efficiency of membrane translocation is unchanged (Fig. 1b).

In these experiments, the organelles were energized for import by the addition of NADH, which in mitochondria of *N. crassa* and yeast is directly channelled into the respiratory chain, resulting in the production of ATP. Does the folding of imported proteins require energy in the form of ATP? We incubated isolated mitochondria in the presence of apyrase, an ATP- and ADP-hydrolysing enzyme from potato. Under these conditions, the urea-denatured precursor is readily translocated into mitochondria. The DHFR component of the imported protein, however, remains sensitive towards digestion with protease, indicating that the refolding reaction is largely inhibited (Fig. 1c). Results were similar when import was performed in the presence of the non-hydrolysable ATP analogues, AMP-PNP and AMP-PCP (AMP-PNP = adenosine-5'-(β , γ -imido)-triphosphate AMP-PCP = adenosine-5'-(β , γ -methylene)-triphosphate) (Fig. 1d). The non-hydrolysable ATP analogues compete with endogenous ATP present in the energized mitochondria. (In contrast to the previous view, membrane translocation of the precursor is not obligatorily coupled with its immediate refolding inside the mitochondria). Folding of the imported protein in the matrix is mediated by an ATP-dependent reaction.

Association with hsp60

Are the folding intermediates associated with a matrix component(s) during the folding reaction? PreSu9-DHFR was imported into mitochondria and digitonin extracts containing 70–90% of the soluble matrix marker fumarate and corresponding amounts of the imported protein were analysed on Sephacryl S-300 sizing columns. Matrix extracts of mitochondria incubated for import for 5 min at 25 °C in the presence of ATP contain the mature-sized fusion protein, which fractionates with a molecular weight corresponding to the monomer (Fig. 2a; i); the DHFR part of the monomeric Su9-DHFR has reached its protease-resistant conformation. About 10% of the protein in

the extract migrates as a high-molecular-weight complex of ~700K. This material represents unfolded, or at least incompletely folded, fusion protein which is sensitive to protease. In extracts of ATP-depleted mitochondria, up to 60% of the protein is recovered in these high-molecular weight fractions (Fig. 2a; ii). This material is highly sensitive to protease (Fig. 4c). No more than 1% of the radiolabel contained in Su9-DHFR is detectable on 17.5% polyacrylamide gels as distinct proteolytic fragments. An antibody directed against denatured DHFR precipitated the protease-sensitive Su9-DHFR as efficiently as the SDS-denatured fusion protein, but did not recognize the folded monomeric form (not shown). Unfolded Su9-DHFR accumulated in the presence of AMP-PNP was also recovered in the high-molecular weight column fractions (Fig. 2a; iii).

It seems that the folding intermediates are probably associated with a complex which participates in the ATP-dependent folding reaction. The digitonin extracts analysed on Sephacryl columns contained ~80% of the total hsp60 present in mitochondria. This hsp60 behaves as a high-molecular weight complex which cofractionates with the incompletely folded intermediates of imported Su9-DHFR (Fig. 2). By contrast with the cofractionating Su9-DHFR, the hsp60 scaffold itself is largely resistant to digestion by protease, although a small fragment of 1–2K is cleaved by proteinase K from either the N- or C-terminus of the hsp60 subunits (Fig. 2a; iv). This does not result in the dissociation of the hsp60 oligomer, which we have demonstrated by re-chromatography of the protease-treated column fractions (not shown).

Authentic mitochondrial proteins such as the Rieske iron sulphur (Fe/S) protein of complex III and the β -subunit of F_1F_0 -ATPase ($F_1\beta$) also form a high-molecular weight complex in the matrix. Fe/S protein imported into apyrase-treated mitochondria is readily extractable with digitonin (Fig. 2b), unlike the protein that is assembled into complex III when imported in the presence of ATP¹⁸. The Fe/S protein contained in the extracts cofractionates exactly with hsp60. Essentially the same result is obtained when $F_1\beta$ is imported: the precursor

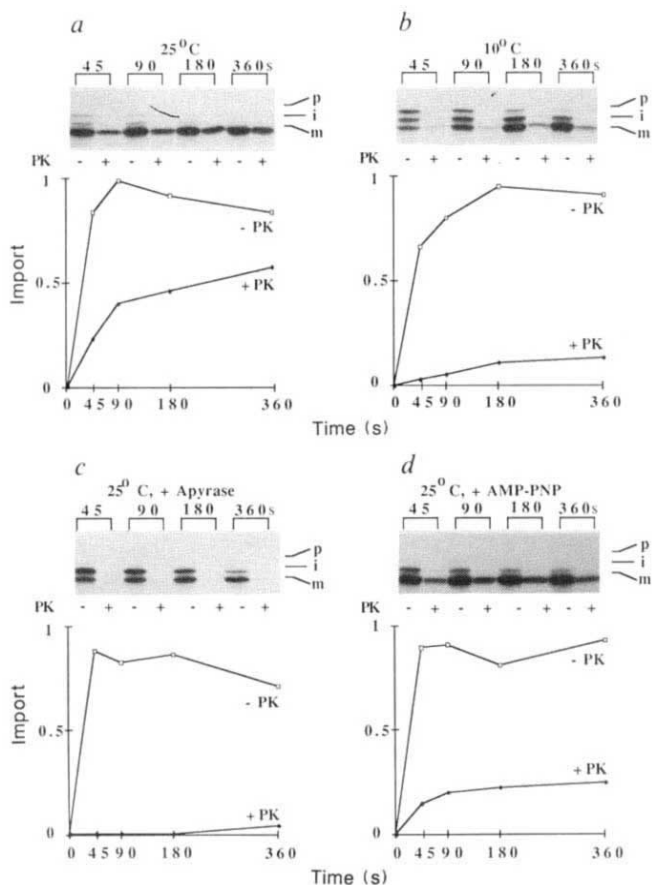


FIG. 1 Import into isolated mitochondria and refolding of Su9-DHFR. Isolated mitochondria were incubated under various conditions in the presence of preSu9-DHFR. Import (arbitrary units) at a, 25 °C; b, 10 °C; c, 25 °C into mitochondria treated with apyrase; d, 25 °C in the presence of 5 mM AMP-PNP. Upper panels, fluorographs of SDS-polyacrylamide gels of digitonin-permeabilized mitochondria previously incubated in the presence (+PK) or in the absence of proteinase K (–PK). p, Precursor; i, intermediate; m, mature-sized form of Su9-DHFR. Lower panels: amount of total imported and protease-resistant Su9-DHFR, as determined by densitometry of the fluorographs.

METHODS. Pre-Su9-DHFR was synthesized in a reticulocyte lysate in the presence of [³⁵S]methionine by transcription/translation of the complementary DNA cloned into expression vector pSP6 (refs 31, 47 and 48). Precursor was precipitated with ammonium sulphate at 66% saturation (30 min at 0 °C) and the precipitate was dissolved in 8 M urea/10 mM Tris, pH 7.5. Denatured precursor was diluted 20–40-fold into import reactions containing 3% BSA, 70 mM KCl, 2.5 mM MgCl₂, 2 mM NADH, 10 mM MOPS (3-[N-morpholino] propanesulphonic acid), pH 7.2, and 0.8 mg ml^{–1} isolated mitochondria of *N. crassa*, which were essentially free of cytosolic and endoplasmic reticulum markers^{18,19}. Before addition of preSu9-DHFR and NADH, reaction c was incubated for 15 min with apyrase (40 U per ml final concentration; Sigma grade VIII)³¹ and reaction d was incubated with 5 mM AMP-PNP. After incubation at 25° (a, c and d) or 10 °C (b) for the times indicated, import reactions were cooled on ice and diluted 5-fold with ice-cold SEM-buffer (0.25 M sucrose, 1 mM EDTA, 10 mM MOPS, pH 7.2) containing 8 μ M antimycin A and 20 μ M oligomycin, then treated with 25 μ g ml^{–1} proteinase K for 10 min at 0 °C. After addition of 1 mM PMSF (phenylmethylsulphonyl fluoride), mitochondria were re-isolated in SEM/100 mM KCl and resuspended at 0.5 mg ml^{–1} in 0.3% digitonin in SEM/100 mM KCl. Half of each reaction was treated with 10 μ g ml^{–1} proteinase K for 10 min at 0 °C. PMSF was added and trichloroacetic acid precipitates analysed by SDS-polyacrylamide gel electrophoresis SDS-PAGE, fluorography and densitometry^{15,19,49}.

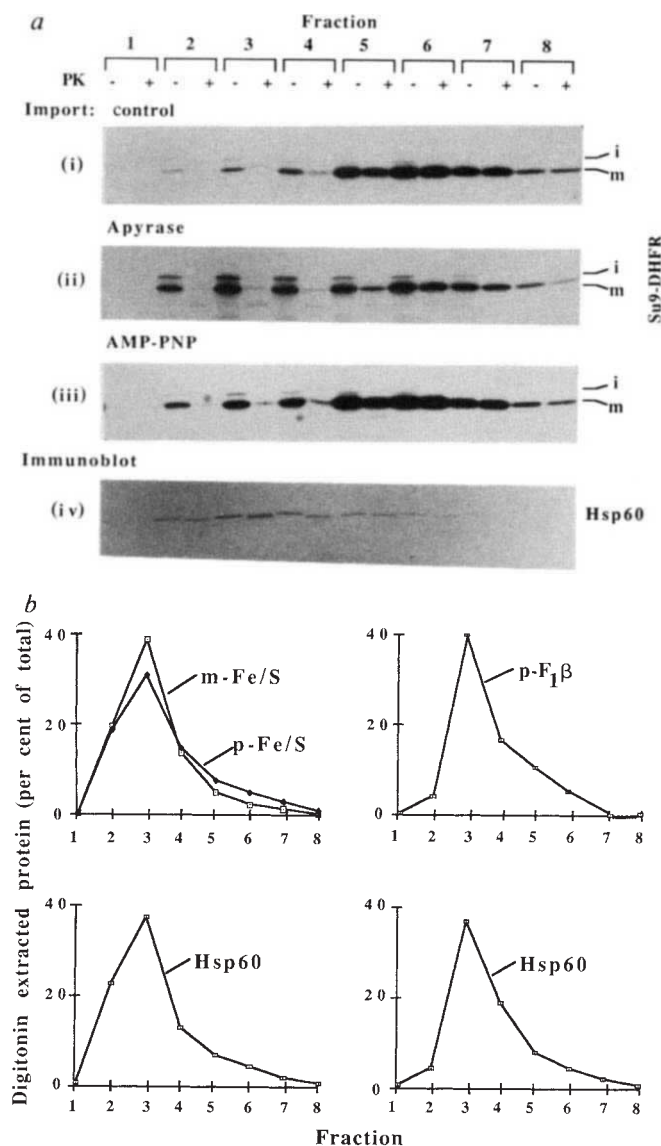


FIG. 2 Gel chromatography of folding intermediates. *a*, Analysis by Sephacryl S-300 chromatography of digitonin extracts prepared from mitochondria which had been incubated for import of preSu9-DHFR: (i) in the presence of ATP; (ii) after depletion of ATP by apyrase; (iii) in the presence of AMP-PNP (iv) Nitrocellulose blot of column fractions decorated with anti-hsp60 antiserum. Fractions of panel (ii) are shown representatively. Column fractions were analysed for total content in Su9-DHFR (–PK) and in protease-resistant Su9-DHFR (+PK). *b*, Gel chromatography of a digitonin extract from apyrase-treated mitochondria which had imported the precursor of Fe/S protein or of $F_1\beta$, respectively. Imported protein and hsp60 are shown as per cent of total protein in digitonin extracts.

METHODS. Import of urea-denatured precursor proteins into isolated mitochondria and protease-treatment of mitochondria are described in Fig. 1. Incubation for import was for 5 min at 25 °C. Matrix was extracted by incubating reisolated mitochondria for 1 min at 0 °C in SEM containing 0.3% digitonin. The concentration of mitochondrial protein was 5 mg ml⁻¹ (ref. 18). After 5-fold dilution with SEM/100 mM KCl, reactions were centrifuged for 10 min at 20,000g. Aliquots of pellets and supernatants were analysed for imported protein, hsp60 and the matrix marker enzyme, fumarate³³. Supernatants were fractionated on 2.5 ml Sephacryl S-300 columns equilibrated with 100 mM NaCl, 10 mM Tris, pH 7.5. The void volume of the column was discarded and 200 µl fractions collected. The peak concentration of the 700K thyroglobulin marker was in fraction 3. Half of each fraction was treated with 10 µg ml⁻¹ proteinase K for 10 min at 0 °C. Trichloroacetic acid precipitates were analysed by SDS-PAGE and fluorography, as well as by immunoblotting⁵⁰ with anti-hsp60 antiserum. Bound antibodies were detected by alkaline phosphatase⁵¹ coupled to IgG directed against rabbit IgG (*a*; iv) or by ¹⁴C-labelled protein A¹⁵ (*b*). Fluorographs were quantified by densitometry.

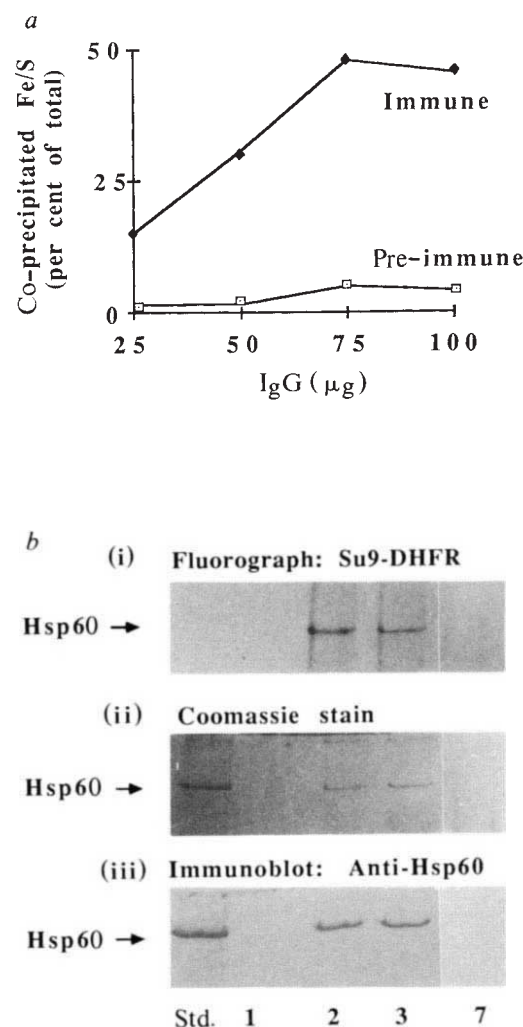


FIG. 3 Physical association of imported proteins with hsp60. *a*, Co-immunoprecipitation of Fe/S protein imported into apyrase-treated mitochondria with anti-hsp60 immunoglobulins. The amount of Fe/S protein (p-Fe/S plus m-Fe/S; compare Fig. 4*b*) precipitated from digitonin extracts are given as per cent of total Fe/S protein imported into mitochondria. The amounts of pre-immune and immune IgG indicated were preabsorbed to protein A-Sepharose. *b*, Non-denaturing polyacrylamide gel electrophoresis of Su9-DHFR-hsp60 complex partially purified by gel chromatography (corresponding to fractions 1–3 in Fig. 2*a*, ii). In addition, the monomeric fusion protein (fraction 7 in Fig. 2) was analysed. i–iii, Fluorograph, Coomassie blue stained and immunoblotted with anti-hsp60 antiserum, respectively. The position of a purified hsp60 standard (Std.) on the gel is indicated.

METHODS. *a*, Fe/S protein was imported into apyrase-treated mitochondria. Digitonin extracts were prepared (see legend to Fig. 2). Aliquots of the extracts corresponding to 25 µg intact mitochondria were added to protein A-Sepharose (PAS) pellets to which 25, 50, 75 and 100 µg preimmune- and hsp60 immunoglobulin had been preabsorbed, respectively¹⁵. Reactions were incubated for 30 min at 4 °C by rotating end-over-end in the presence of 1 mM PMSF and 30 µg ml⁻¹ protease inhibitor from *N. crassa*³³. PAS beads were pelleted and washed twice with SEM/KCl buffer and once with 30 mM Tris, pH 7.4. Wash solutions also contained protease inhibitors. PAS-immunoglobulin-antigen complexes were dissociated in SDS-containing buffer¹⁵ and analysed by SDS-PAGE, fluorography and densitometry. The total amount of imported Fe/S protein contained in the extract was determined by immunoprecipitation with saturating amounts of anti-Fe/S immunoglobulin. *b*, Digitonin extracts of apyrase-treated mitochondria which had imported preSu9-DHFR were fractionated by gel chromatography. 50 µl aliquots of fractions 1–3 and of fraction 7 (see legend to Fig. 2) were loaded twice on the same 4–20% non-denaturing polyacrylamide gel³⁵. Isolated hsp60 (1 µg) was analysed as standard. After electrophoresis, one half of the gel was blotted onto nitrocellulose and decorated with anti-hsp60 antiserum; the other half was stained with Coomassie blue and analysed by fluorography.

form accumulates in the apyrase-treated mitochondria and is detected as high-molecular weight complex (Fig. 2*b*).

We then demonstrated that the proteins imported at low levels of ATP are physically associated with hsp60. Fe/S protein was imported into apyrase-treated mitochondria, which were then extracted with digitonin. Monospecific immunoglobulins directed against hsp60 could co-immunoprecipitate up to 45% of the total Fe/S protein contained in the extracts (Fig. 3*a*). Comparable results were obtained with coprecipitation of Su9-DHFR (not shown), demonstrating that the column cofractionation of the imported proteins with hsp60 reflects a stable interaction between these components. Analysis of the high-molecular weight column fractions containing Su9-DHFR by non-denaturing gel electrophoresis³⁵ also indicated such an association. The protease-sensitive Su9-DHFR comigrated exactly with the hsp60 complex (Fig. 3*b*); between 50–70% of Su9-DHFR and hsp60 contained in the column fractions was recovered on the non-denaturing gel. The monomeric folded fusion protein was not seen as a distinct band.

Together our results indicate that proteins imported into the mitochondrial matrix transiently associate as incompletely folded intermediates in a high-molecular weight assembly com-

plex before attaining their final, folded structure. Hsp60 is identified as a major constituent of this complex.

ATP-dependent folding and release

We investigated the ATP-requiring step in the sequence of reactions involved in folding imported proteins by adding back ATP to the hsp60-bound folding intermediates to see if they could be chased into the folded species. PreSu9-DHFR was imported into ATP-depleted mitochondria and matrix extracts were incubated with or without Mg^{2+} ATP. At low levels of ATP, about 60% of the imported fusion protein contained in the matrix extract remains protease-sensitive and cofractionates with hsp60 (Fig. 4*a*). In contrast, addition of ATP causes folding into the protease-resistant conformation and the release of Su9-DHFR from hsp60. The mature-sized protein fractionates as the monomer (Fig. 4*a*). The fractionation of hsp60 is unchanged. Apparently, the folding intermediates bound to hsp60 in the absence of ATP are productive intermediates on the authentic import pathway.

We also demonstrated this for the Fe/S protein associated with hsp60 in apyrase-treated mitochondria (Fig. 4*b*). Readdition of ATP to the intact mitochondria causes the release of

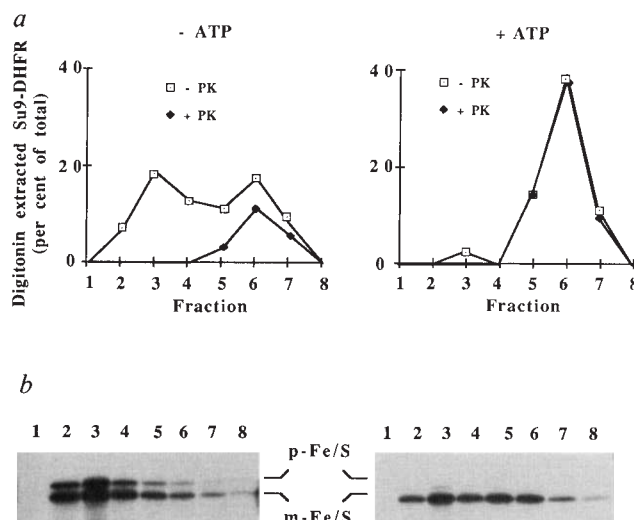
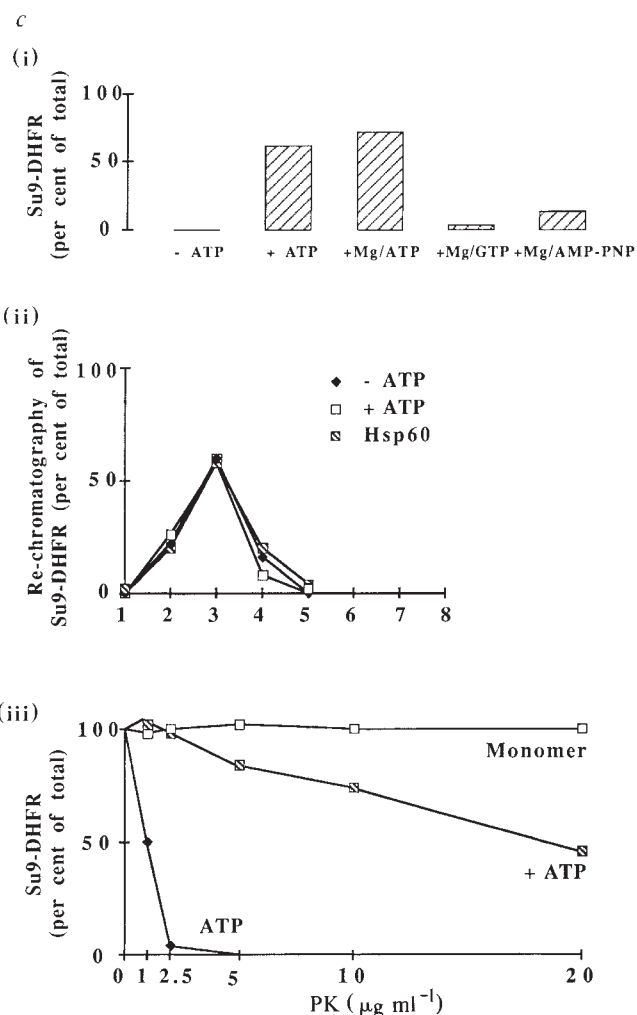


FIG. 4 Folding of imported proteins and their release from hsp60. *a*, ATP-dependent folding and release of imported Su9-DHFR. Before gel chromatography, digitonin extracts of mitochondria were incubated for 15 min at 25 °C in the absence (-ATP) or presence of 5 mM Mg^{2+} ATP (+ATP). Half of each column fraction was treated with proteinase K (+PK). The amount of total Su9-DHFR and of protease-resistant Su9-DHFR is given in per cent of total protein in digitonin extracts. *b*, ATP-dependent release of imported Fe/S protein from hsp60. After import into apyrase-treated mitochondria, the re-isolated organelles were incubated for 15 min at 25 °C in the absence (left panel) or presence of 5 mM Mg^{2+} ATP (right panel). Digitonin extracts were fractionated by gel chromatography. *c*, NTP-dependence of folding of Su9-DHFR: (i) Su9-DHFR-hsp60 complex partially purified by gel chromatography (pooled fractions 2–4, see Fig. 2, ii) was analysed after incubation for 15 min at 25 °C in the absence or presence of 5 mM NTPs as indicated. The amount of protease-resistant Su9-DHFR produced is given in per cent of total Su9-DHFR analysed. (ii) Re-chromatography of Su9-DHFR-hsp60 complex after incubation in the absence or presence of ATP. Fractionation of hsp60 in the +ATP reaction is shown representatively. Amounts are given as per cent of total protein analysed. (iii) Protease-resistance of Su9-DHFR contained in the complex with hsp60 after incubation with or without ATP, and of the monomer Su9-DHFR obtained in *a*. Protein resistant to proteinase K at the concentrations indicated is shown as per cent of total protein. Equal amounts of labelled Su9-DHFR were analysed. METHODS. Import of precursor proteins into isolated mitochondria, preparation of digitonin extracts, gel chromatography and analysis of column fractions are described in the legend to Fig. 2. *c*, Column fractions 2–4 (see Fig. 2, ii) containing protease-sensitive Su9-DHFR exclusively, were pooled



and divided into five 120 μl aliquots. Before incubation for 15 min at 25 °C, these reactions were made 5 mM in Na_2ATP , $Mg^{2+}ATP$, $Mg^{2+}GTP$ or $Mg^{2+}AMP-PNP$ using 100-fold-concentrated stock solutions in water adjusted to pH 7. Half of each reaction was treated with 10 $\mu g\ ml^{-1}$ proteinase K for 10 min at 0 °C. Treatment with proteinase K for titration of protease resistance was under the same conditions. Trichloroacetic acid precipitates were analysed by SDS-PAGE, fluorography and densitometry.

60% of the hsp60-bound protein. Precursor is processed to m-Fe/S. Half of the released m-Fe/S is no longer extractable by digitonin and is recovered in the membrane fraction: probably this membrane-associated m-Fe/S is assembled into complex III¹⁸. The digitonin-extractable portion of the m-Fe/S that has been released from hsp60 migrates with a molecular weight of ~70K, which is higher than that expected for the monomeric Fe/S protein (25K). This m-Fe/S could represent protein not yet tightly assembled into complex III, or another matrix-localized intermediate on the assembly pathway. In contrast to the observations involving Su9-DHFR, ATP-dependent release of the Fe/S protein from hsp60 can only be demonstrated with intact mitochondria, and not with the matrix extracts containing the Fe/S-hsp60 complex. As the Fe/S protein has to be re-exported across the inner membrane, the inner membrane itself or a membrane-associated factor(s) could be required for its release from hsp60.

We investigated the nucleoside triphosphate specificity of the folding reaction using the Su9-DHFR-hsp60 complex from ATP-depleted mitochondria after its partial purification by gel chromatography. Aliquots were incubated for 15 min at 25 °C with and without NTPs (Fig. 4c; i). Addition of Mg²⁺ATP was most effective in promoting folding of the DHFR moiety, when about 60% of the total fusion protein remained resistant to proteinase K at 10 µg ml⁻¹. GTP and AMP-PNP were ineffective in promoting folding. In contrast to our results using total matrix extracts (Fig. 4a), the protease-resistant fusion protein still cofractionated with the hsp60 complex (Fig. 4c; ii) and could be co-immunoprecipitated by anti-hsp60 antibodies as efficiently as the unfolded fusion protein present in the absence of ATP (not shown). Titration with increasing concentrations of proteinase K revealed that this Su9-DHFR is less resistant to digestion than the monomeric fusion protein (Fig. 4c; iii). Several protease-resistant fragments accounting for ~10% of the radiolabel contained in Su9-DHFR were detected on SDS-polyacrylamide gels near the gel front (not shown). This indicates that ATP-mediated folding of the DHFR moiety is occurring at the surface of the hsp60 complex. It is possible that an hsp60 function necessary for release of the associated polypeptides is inactivated during gel chromatography or, more likely, that an additional component(s) of the mitochondrial matrix which does not cofractionate with hsp60 is necessary for the complete sequence of reactions leading to the folded monomeric protein.

NEM-sensitivity of folding

Is the folding of Su9-DHFR under physiological import conditions, that is, in the presence of intramitochondrial ATP, also dependent on protein factors? Using the membrane-permeant alkylating agent, N-ethylmaleimide (NEM), we found that modification of mitochondria with 1–5 mM NEM before import in the presence of ATP inhibited the refolding of imported Su9-DHFR into a protease-resistant conformation (Fig. 5a). Translocation of the urea-denatured precursor into the mitochondria and proteolytic processing were unaffected. NEM treatment does not cause depletion of ATP in the matrix compartment (see legend to Fig. 5). Interestingly, after short import incubation (2–5 min), Su9-DHFR can be extracted by digitonin and is associated with hsp60 (Fig. 5b), whereas in control reactions the protein is folded and exists as the monomer (Fig. 2a). But after incubation of the NEM-treated mitochondria for 15 min, most of the imported protein is no longer extractable and is recovered in the membrane fraction. This Su9-DHFR was still very sensitive to protease. Extractability of hsp60 was unchanged, indicating that Su9-DHFR had fallen off the hsp60 complex and was forming incompletely folded aggregates. After combined treatment with NEM and aprotinin, readdition of ATP causes a shift of the protease-sensitive Su9-DHFR from the extractable complex with hsp60 into the membrane fraction (not shown). The aggregated fusion protein is slightly more protease-

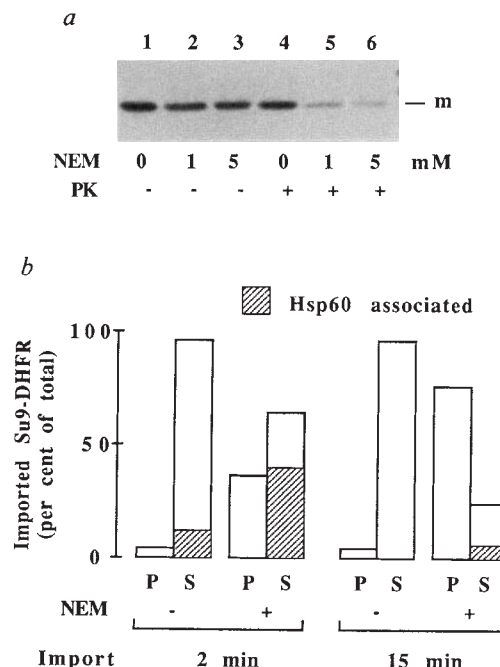


FIG. 5 Inhibition of folding in NEM-treated mitochondria. *a*, Import of preSu9-DHFR for 15 min at 25 °C into isolated mitochondria pretreated with NEM at the concentrations indicated. Mitochondria were permeabilized with digitonin and were incubated in the absence (lanes 1–3) or in the presence (lanes 4–6) of proteinase K. *b*, Extractability by digitonin of Su9-DHFR imported for 2 min or 15 min into control mitochondria and into mitochondria pretreated with 5 mM NEM. Extracts were separated into pellets (P) and supernatants (S). Supernatants were analysed by gel chromatography. Column fractions 3 and 4 containing the highest concentrations of hsp60-complex were pooled. Imported Su9-DHFR is given as per cent of total imported protein. The shaded area of the columns marked supernatant represents Su9-DHFR recovered in hsp60-containing column fractions.

METHODS. Mitochondria were incubated for 10 min at 25 °C at a protein concentration of 2 mg ml⁻¹ in SEM-buffer containing 0, 1 or 5 mM NEM. Dithiothreitol was added to a concentration 5-fold higher than NEM and incubation continued for 5 min at 25 °C. Mitochondria reisolated by centrifugation were resuspended in SEM and used for import of Su9-DHFR as described in Fig. 1. Import reactions contained 2 mM Mg²⁺ATP. Digitonin treatment, preparation of digitonin extracts and analysis on Sephacryl S-300 columns are described in the legends to Figs 1 and 2. Oligomycin-sensitive ATP hydrolysis of NEM-treated mitochondria was assayed: mitochondria were incubated as for import in the presence of 1 mM Mg²⁺ATP for various times at 25 °C either with or without 20 µM oligomycin and ATP in the mitochondrial supernatant was determined by coupled enzyme assay⁵². Compared with controls, NEM-treated mitochondria internalized and hydrolysed identical amounts of ATP.

resistant than the polypeptides bound to hsp60 in the absence of ATP. Formation of insoluble aggregates of imported proteins, including a defect in the refolding of Su9-DHFR, was also observed in mitochondria of the hsp60-deficient yeast mutant, mif4 (not shown).

We conclude that folding of the imported protein must be protein 'catalysed': folding at the surface of hsp60 and ATP-dependent release are two essential but separable steps necessary to achieve the final conformation. It remains to be tested whether hsp60 itself is the target of the NEM-effect. We have detected a reaction of hsp60 with ¹⁴C-labelled NEM under conditions which inhibit folding (not shown).

Discussion

We have described here the surprising discovery that proteins imported from the cytosol into mitochondria do not refold spontaneously once translocation across the mitochondrial membranes is complete. The stress protein hsp60 residing in the mitochondrial matrix has been identified as a major component of an NEM-sensitive device involved in the folding of imported

proteins, which acts in conjunction with ATP. In the absence of ATP, newly imported proteins form a soluble complex with hsp60. Newly imported protein is associated in a loosely folded conformation, which is very sensitive to protease, with the surface of the hsp60 scaffold. ATP hydrolysis allows folding and release from hsp60, but our present data do not prove that this ATP effect(s) is exerted by hsp60 itself. On the other hand, ATP-dependent folding of the associated polypeptide occurs at the surface of hsp60 before release. It remains to be determined whether the release of the partially folded polypeptide from hsp60 has a direct ATP requirement, or whether it is merely a consequence of the preceding ATP-dependent folding. One or more additional component(s) might be involved. Mitochondria could contain an equivalent to the groES protein, which in *E. coli* seems to cooperate with groEL in various assembly functions by an unknown mechanism³⁶⁻³⁸.

How does hsp60 fulfil its function? A parallel study has shown that proteins traverse the mitochondrial membranes in an extended conformation (unpublished results). We propose that polypeptides entering the mitochondrial matrix have a conformation similar to nascent chains emerging from the ribosome, and that it is these unfolded polypeptides that interact with hsp60. This may even occur during translocation. Our results with the DHFR-fusion protein indicate that hsp60 is not specific for the structural motifs present in only mitochondrial proteins. Notably, urea-denatured DHFR, when part of a fusion protein, can refold spontaneously in appropriate buffer solutions³⁴, but, this does not happen after translocation across the mitochondrial membranes. The mitochondrial matrix is effectively a highly concentrated protein solution, which does not correspond to conditions frequently used in refolding experiments *in vitro*. It may be an essential function of hsp60 to capture the unfolded polypeptides entering the matrix and so prevent the formation of misfolded proteins. Cytosolic precursors of mitochondrial and secretory proteins interacting with 70K heat-shock proteins^{3,4} have been proposed to bind through hydrophobic regions exposed in the newly synthesized, incompletely folded polypeptides^{6,7,39,40}. The complex formed between hsp60 and imported polypeptides is only slightly destabilized by non-ionic detergents such as Triton X-100 (results not shown); this could indicate that here the interaction is not predominantly hydrophobic in nature.

The role of ATP in hsp60-mediated folding and release of the polypeptide is unclear. Although the chaperonins, including the

mitochondrial hsp60, are structurally unrelated to the 70K heat-shock proteins, both groups of components exhibit weak ATPase activity³⁹⁻⁴³. It has been suggested that ATP hydrolysis causes a conformational change in 70K heat-shock proteins, which is transferred to the bound polypeptide^{6,7,39,40,44}. An additional soluble factor seems to participate in these reactions^{4,5,9}. So far, the function of the 70K proteins has been evident in stabilization of a loose, translocation-competent conformation of cytosolic precursor proteins, or the disruption of protein aggregates. In contrast, the mitochondrial hsp60 is essential for protein folding. ATP hydrolysis by hsp60 or by another component could loosen the association of the unfolded polypeptides with the hsp60 scaffold, and thus allow for their ordered, domain-wise folding at the surface of hsp60. We describe this type of reaction as 'protein-catalysed protein folding'; it may be slower than the 'spontaneous folding' that occurs in the mitochondrion in the absence of the functional folding 'catalyst' resulting in misfolded protein aggregates. Hsp60 and the homologous groEL protein of *E. coli* may have similar functions in folding nascent polypeptide chains synthesized in the mitochondrial matrix or in the bacterial cytosol. It is possible that protein folding in the eukaryotic cell cytosol or in other membrane compartments such as the endoplasmic reticulum, could also be protein-mediated.

The chloroplast Rubisco-binding protein and hsp60 were originally defined as components assisting in oligomeric protein assembly^{20,21,25,28,29}. In the light of our results, these chaperoning reactions probably represent only a part of their function. The actual assembly reactions could occur spontaneously after folding of the subunits and their release from hsp60. Alternatively, complementary surfaces of subunits might be exposed only while the partially folded polypeptides are still in an assembly complex with hsp60. Likewise proteins such as the Rieske Fe/S protein or cytochrome *b*₂, which have to be re-exported from the matrix to the intermembrane space^{18,19}, could be shuttled to their respective protein export centres while associated with hsp60. A function for groEL in the stabilization of proteins for export across the bacterial plasma membrane has been suggested⁴⁵. Other factors in the mitochondrial matrix, such as the recently described 70K protein SSC1⁴⁶, could also be involved in keeping proteins in a translocation-competent conformation.

Our findings concerning the function of hsp60 should have application in the renaturation by folding catalysts of proteins obtained by overexpression in bacteria, raising interesting possibilities for biotechnologists in the future. □

Received 7 July; accepted 7 August 1989.

- Attardi, G. & Schatz, G. A. *Rev. Cell Biol.* **4**, 289-333 (1988).
- Hartl, F.-U., Pfanner, N., Nicholson, D. & Neupert, W. *Biochim. biophys. Acta* **988**, 1-45 (1989).
- Deshaies, R. J., Koch, B. D., Werner Washburne, M., Craig, E. & Schekman R. *Nature* **332**, 800-805 (1988).
- Chirico, W. J., Waters, M. G. & Blobel, G. *Nature* **332**, 805-810 (1988).
- Zimmermann, R., Sagstetter, M., Lewis, M. J. & Pelham, H. R. B. *EMBO J.* **7**, 2875-2880 (1988).
- Pelham, H. R. B. *Cell* **46**, 959-961 (1986).
- Pelham, H. R. B. *Nature* **332**, 776-777 (1988).
- Rothman, J. E. & Kornberg, R. D. *Nature* **322**, 209-210 (1986).
- Murakami, H., Pain, D. & Blobel, G. *J. Cell Biol.* **107**, 2051-2057 (1988).
- Schleyer, M. & Neupert, W. *Cell* **43**, 339-350 (1985).
- Schwaiger, M., Herzog, V. & Neupert W. *J. Cell Biol.* **105**, 235-246 (1987).
- Vestweber, D. & Schatz, G. *J. Cell Biol.* **107**, 2037-2043 (1988).
- Rassow, J. *et al. J. Cell Biol.*, in the press.
- Hartl, F.-U. & Neupert, W. in *Structural and Organizational Aspects of Metabolic Regulation* UCLA Symposium on Molecular and Cellular Biology, Vol. 134 (eds Srere, P. *et al.*) (Liss, New York, in the press).
- Hawltischek, G. *et al. Cell* **53**, 795-806 (1988).
- Witte, C., Jensen, R. E., Yaffe, M. P. & Schatz, G. *EMBO J.* **7**, 1439-1447 (1988).
- Pollock, R. A. *et al. EMBO J.* **7**, 3493-3500 (1988).
- Hartl, F.-U., Schmidt, B., Wachter, E. & Neupert, W. *Cell* **47**, 939-951 (1986).
- Hartl, F.-U., Ostermann, J., Guillard, B. & Neupert W. *Cell* **51**, 1027-1037 (1987).
- Cheng, M. Y. *et al. Nature* **337**, 620-625 (1989).
- Reading, D. S., Hallberg, R. L. & Myers, A. M. *Nature* **337**, 655-659 (1989).
- Eilers, M. & Schatz, G. *Cell* **52**, 481-483 (1988).
- McMullin, T. W. & Hallberg, R. L. *Molec. cell. Biol.* **8**, 371-380 (1988).
- Hutchinson, E. G., Tichelaar, W., Hofhaus, G., Weiss, H. & Leonhard, K. R. *EMBO J.* **8**, 1485-1490 (1989).
- Hemmingsen, S. M. *et al. Nature* **333**, 330-334 (1988).
- Waldinger, D., Eckerskorn, C., Lottspeich, F. & Cleve, H. *Biol. Chem. Hoppe-Seyler* **369**, 1185-1189 (1988).
- Jindal, S., Dudani, A. K., Singh, B., Harley, C. B. & Gupta, R. S. *Molec. cell. Biol.* **9**, 2279-2283 (1989).

- Ellis, R. J. *Nature* **328**, 378-379 (1987).
- Ellis, R. J. & van der Vies, S. M. *Photosynthesis Res.* **16**, 101-115 (1988).
- Eilers, M. & Schatz, G. *Nature* **322**, 228-232 (1986).
- Pfanner, N., Tropschug, M. & Neupert, W. *Cell* **49**, 815-823 (1987).
- Viebrock, A., Perz, A. & Sebald, W. *EMBO J.* **1**, 565-571 (1982).
- Schmidt, B., Wachter, E., Sebald, W. & Neupert, W. *Eur. J. Biochem.* **144**, 581-588 (1984).
- Eilers, M., Hwang, S. & Schatz, G. *EMBO J.* **7**, 1139-1145 (1988).
- Musgrove, J. E., Johnson, R. A. & Ellis, R. J. *Eur. J. Biochem.* **163**, 529-534 (1987).
- Tilly, K., Murialdo, H. & Georgopoulos, C. P. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1629-1633 (1981).
- Chandrasekhar, G. N., Tilly, K., Woolford, C., Hendrix, R., Georgopoulos, C. *J. Biol. Chem.* **261**, 12414-12419 (1986).
- Golubinioff, P., Gatenby, A. A. & Lorimer, G. H. *Nature* **337**, 44-47 (1989).
- Munro, S. & Pelham, H. R. B. *Cell* **46**, 291-300 (1986).
- Kassenbrock, C. K., Garcia, P. D., Walter, P. & Kelly, R. D. *Nature* **333**, 90-93 (1988).
- Hendrix, R. W. *J. Molec. Biol.* **129**, 375-392 (1979).
- Chandari, P., Cannon, S., Hubbs, A. & Roy, H. *Abstr. 16th Ann. UCLA Symp. J. Cell Biochem. suppl.* **118**, 50 (1987).
- Chapell, T. G. *et al. Cell* **45**, 3-13 (1986).
- Kassenbrock, C. K. & Kelly, R. B. *EMBO J.* **8**, 1461-1468 (1989).
- Bochkareva, E. S., Lissin, N. M. & Girshovich, A. S. *Nature* **336**, 254-257 (1988).
- Craig, E. A., Kramer, J. & Kosic-Smithers, J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4156-4160 (1987).
- Pelham, H. R. B. & Jackson, R. J. *Eur. J. Biochem.* **67**, 247-256 (1976).
- Krieg, P. A. & Melton, D. A. *Nucleic Acids Res.* **12**, 7057-7070 (1984).
- Laemmli, U. K. *Nature* **227**, 680-685 (1970).
- Burnette, W. H. *Analyt. Biochem.* **112**, 195-203 (1981).
- Blake, M. S., Johurta, K. H., Russel-Jones, G. J. & Gotschlick, E. C. *Analyt. Biochem.* **136**, 175-179 (1984).
- Bergmeyer, H. U. in *Methods of Enzymatic Analysis* Vol. VII, 346-357 (Springer, Berlin, Heidelberg & New York, 1985).

ACKNOWLEDGEMENTS. We thank Dr H. Pelham for antibody against BiP-protein. This work was supported by the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie, by the Genzentrum München, by the NIH, and by the Bernard and Jennie M. Nelson Fund.

The asymmetric monopole and non-newtonian forces

J. W. Barrett

Department of Physics, University of Newcastle upon Tyne,
Newcastle upon Tyne NE1 7RU, UK

A SOLID spherical object of constant density is a monopole for newtonian gravity. It seems perfectly plausible that monopoles of other shapes ('asymmetric monopoles') exist, and indeed, I have found approximate numerical solutions to this problem. Such asymmetric monopoles are interesting for the detection of deviations from Newton's inverse-square law. Although the newtonian field is monopolar, the field as a result of any deviation will not be. This, in principle, allows one to separate the deviation from the much larger newtonian field, which would normally swamp it on a laboratory scale. A laboratory experiment to detect the proposed 'fifth force' may be feasible.

Figure 1 shows the cross-section through a rotationally symmetric object of constant density that has a hollow interior. The object is created by rotating the figure through the horizontal axis (not shown). The gravitational potential outside the object (referred to as the 'apple' because of its apple-shaped hollow) is $-GM/r$ (where G is the gravitational constant, M is the mass and r is the distance), plus a small error field. The shape is not the only one that may be used to generate an asymmetric monopole, but it is the most suitable from the range that I considered.

In an experiment, one would exploit the fact that the newtonian field is monopolar by rotating the apple about a different axis, for example the vertical axis drawn on Fig. 1, which goes through the centre of mass. The field that results from deviations from Newton's law would then oscillate, and could be observed by using the same apparatus used to detect gravity waves. The use of gravity-wave detectors for this purpose has been suggested or tried in various different ways (see, for example, refs 1, 2 and A. Rüdiger and G. W. Gibbons, personal communication). I assume that the deviation is of 'fifth force' type, as has been suggested by some experiments summarized in refs 3 and 4. The potential of a mass point of mass M has an added term $-\alpha GM/re^{(-r/\lambda)}$ where the dimensionless strength $|\alpha|$ is 0.01–0.05 and the range of the interaction λ is 15–500 m. There is a non-zero dipole field for the fifth force, because the multipole expansion for the fifth force is different from the multipole expansion for newtonian gravity⁵.

I assume, as an example, that the apple has a diameter of 1 m and is made from aluminium; its mass is therefore 650 kg.

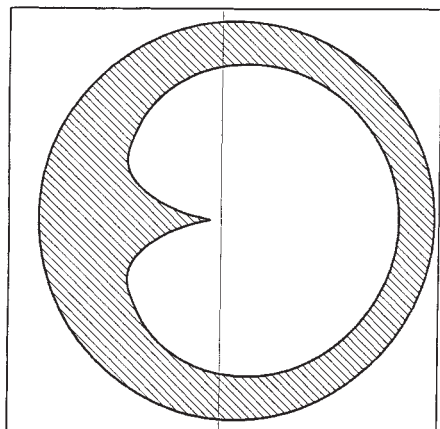


FIG. 1 An asymmetric monopole is obtained by rotating the figure about the vertical axis, which passes through the centre of mass.

It may be that another metal is more suitable. If a freely suspended test mass is placed at a distance r from the centre of mass of the apple on the horizontal equatorial plane, while the apple is rotated about the vertical axis with frequency f , the test mass will respond to the fifth force with an oscillating radial displacement S of r.m.s. amplitude:

$$\frac{\alpha}{r^3 \lambda^2 f^2} \times (9 \times 10^{-13} \text{ m}^6 \text{ s}^{-2})$$

This formula is a good approximation as long as $r \ll \lambda$. If $r \sim \lambda$, then the formula is somewhat more complicated than this.

The key to doing such an experiment, as far as the detector is concerned, is to achieve good sensitivity at a low frequency, where the isolation from seismic noise is greatest. An advantage is that one can run the experiment for a long time and use a very narrow bandwidth response. With today's laser interferometers; a narrow-bandwidth displacement sensitivity of 10^{-21} m is achievable⁶. If this can be achieved at a frequency $f = 10$ Hz, as may be possible with the new generation of isolating pendula^{7,8}, the region $\alpha/\lambda^2 > 10^{-6} \text{ m}^{-2}$ becomes accessible.

There are other constraints on the achievable accuracy from the practical aspect of making the source mass. The shapes of the inside and outside contours are given by the formula $a(\theta) = a_0 + a_1 P_1(\cos \theta) + a_2 P_2(\cos \theta) + \dots + a_{12} P_{12}(\cos \theta)$, where the P s are Legendre polynomials and the coefficients a_n are given in Table 1. The newtonian error field, referred to earlier, will depend on the accuracy to which the shape is made. This error field may interfere with the detection of the fifth force. I calculated the error that would result if the shape was formed to within $1 \mu\text{m}$, that is, an uncertainty of $\pm 0.5 \mu\text{m}$. The error results in small (newtonian) dipole and multipole moments, which will give an oscillating field when the apple is rotated. The relevant $m = 1$ components of the dipole and quadrupole moments are related to the force and couple on the rotating body, and the effects can be measured and controlled. The dipole can be eliminated by shifting the axis of rotation, and the quadrupole contribution is decreased because the harmonic of fundamental frequency vanishes in the equatorial plane. The first moment that cannot be controlled dynamically is the octupole moment. Thus the newtonian error field falls off at least as fast as the octupole moment (r^{-5}) for the displacement measurement S , and one can distinguish a fifth-force dipole from newtonian effects if the results of two measurements at different distances differ by a ratio less than the fifth power of the distance ratio. Table 2 gives the typical errors for a shape formed to within $1 \mu\text{m}$, together with the fifth-force signal strength. The effect on the test object is presented in two different ways—the displacement and the strain. These are multiplied by f , λ and α as shown to give invariant quantities. The quantity h is the usual quantity⁹ of twice the r.m.s. fractional change in length for a small object, with radial alignment. At this level of accuracy; it seems that one can explore the region $\alpha/\lambda^2 > 10^{-5} \text{ m}^{-2}$.

TABLE 1 The outer and inner contours

Coefficient	Outer contour	Inner contour
a_0	0.4682370	0.3482169
a_1	0.0347933	0.0969634
a_2	-0.0034762	-0.0450812
a_3	0.0005227	0.0346095
a_4	-0.0000911	-0.0304927
a_5	0.0000171	0.0276200
a_6	-0.0000034	-0.0245809
a_7	0.0000007	0.0209728
a_8	-0.0000001	-0.0168116
a_9	0	0.0123419
a_{10}	0	-0.0079524
a_{11}	0	0.0041198
a_{12}	0	-0.0013419

TABLE 2 Newtonian error and signal strength

Distance r (m)	Maximum 1- μm shape error		Fifth force	
	Displacement $f^2 S$ (m s^{-2})	Strain $f^2 h$ (s^{-2})	Displacement $f^2 S \lambda^2 / \alpha$ ($\text{m}^3 \text{s}^{-2}$)	Strain $f^2 h \lambda^2 / \alpha$ ($\text{m}^2 \text{s}^{-2}$)
2	4×10^{-18}	2×10^{-17}	1.1×10^{-13}	3×10^{-13}
4	1.3×10^{-19}	3×10^{-19}	1.4×10^{-14}	2×10^{-14}
8	4×10^{-21}	5×10^{-21}	1.8×10^{-15}	1.3×10^{-15}

The results detailed here show that it may be feasible to use the asymmetric monopole to test the fifth-force hypothesis, although the experiment is obviously not easy, and practical problems remain to be examined in detail. For example, there are small elastic distortions in a rotating body as a result of centrifugal forces. The shape defined by Table 1 is the shape that must be achieved at the final rotation speed. A numerical calculation would give the correct initial shape for a given material. Advantages of the experiment are that the experiment is a null one and one can adjust the shape to find the null, and that only octupole-patterned deformations give rise to error signals as large as in Table 2. Also, the parameters I have chosen (size, distance, tolerance and material) are nominal. There is also the possibility that a refinement of this idea, or an incorporation of the mathematical idea into other experimental techniques (such as torsion balances) may prove effective. I note that the experiment would not depend on any variation of the fifth force with chemical composition, as do many of the recent experiments. $\square$

Received 8 May; accepted 31 July 1989.

1. Sinsky, J. A. *Phys. Rev.* **167**, 1145–1151 (1968).
2. Hyakawa, K. & Kuroda, K. *Phys. Rev. D* **32**, 342–346 (1985).
3. Stubbs, C. W. *Nature* **338**, 301–302 (1989).
4. Stacey, F. D. *et al. Rev. Mod. Phys.* **59**, 157–174 (1987).
5. Duff, G. F. D. & Naylor, D. *Differential Equations of Applied Mathematics* 356 (Wiley, New York, 1966).
6. Ward, H. *et al. Proc. Int. Symp. Expl. Gravitational Phys.* (eds Michelson, P. F., Hu, E.-L. and Pizzella, G.) 322–327 (World Scientific, Singapore, 1988).
7. Del Fabbro, R. *et al. Phys. Lett. A* **124**, 253–257 (1987).
8. Del Fabbro, R. *et al. Phys. Lett. A* **132**, 237–240 (1987).
9. Thorne, K. S. in *300 Years of Gravitation* (eds Hawking, S. W. & Israel, W.), 330–458 (Cambridge University Press, 1987).

ACKNOWLEDGEMENTS. I thank G. Gibbons for discussions about the fifth force, J. Hough for discussions about gravity-wave detectors and M. Hindmarsh for practical help.

CO₂ and climate: a missing feedback?

J. F. B. Mitchell, C. A. Senior & W. J. Ingram

Meteorological Office, London Road, Bracknell RG12 2SZ, UK

THE potential of changes in cloud properties to modulate climate perturbations is well known^{1,2}. Changes in cloud amount and cloud height³, cloud radiative properties^{4–6} and cloud condensation nuclei⁷ (CCN) have all been recognized as possible sources of climate feedbacks. Here we report results of simulations that indicate that the changes of state of cloud water may provide a substantial negative feedback on climate. The feedback is concentrated in mid-latitudes and affects both the magnitude and distribution of the climate change expected from increases in 'greenhouse' gases. Improved measurements and parameterizations of cloud processes are needed to quantify this process.

The model is an 11-level atmospheric general-circulation model on a $5^\circ \times 7.5^\circ$ (latitude $\times$ longitude) horizontal grid coupled to a 50-m mixed-layer ocean with a prescribed seasonally and geographically varying oceanic heat convergence to represent heat advection by ocean currents⁸. In the standard

model⁸ cloud cover is a function of relative humidity. A second version of the model was constructed with an explicit cloud-water variable for all but deep convective cloud⁹. A balance is maintained between cloud water content and water vapour, which is altered by dynamical, convective and boundary-layer processes. Cloud water content is depleted by precipitation at a rate dependent on the phase of the cloud water, which is a function of the local temperature of the layer. For water cloud the rate of conversion to precipitation increases as a function of the in-cloud water content and is marginally enhanced in the presence of precipitation from higher layers. The transition from water cloud to ice cloud occurs between -15°C and 0°C as a smoothly varying function of temperature. Ice-cloud particles are assumed to start falling as soon as they form, with a fixed fall speed of 1 m s^{-1} . The rate of precipitation of ice is, on average, an order of magnitude greater than that of water even in the presence of precipitation from the layer above. Cloud radiative properties were prescribed. Note that although the revised cloud scheme is more detailed, it is not necessarily more accurate than the less sophisticated scheme. Our work highlights how different parameterization schemes can affect the sensitivity of the climate to CO₂.

Two simulations, with CO₂ concentrations of 320 and 640 p.p.m., were performed for each version of the model. The results shown here are averaged over five years after the simulations reached equilibrium. On replacing the relative-humidity cloud scheme⁸ (henceforth referred to as RH) with the model based on cloud water content⁹ (CW), the global annual average surface warming is reduced from 5.2 K to 2.7 K. The largest fractional reduction in the warming occurs in higher latitudes (Fig. 1a) where the large decreases in total cloud cover with RH have been replaced by much smaller reductions or slight increases (Fig. 1b).

With RH, reductions in cloud cover are found throughout the depth of the mid-latitude troposphere (Fig. 2a), as in earlier studies³. With CW, the main differences in response occur near the freezing level (Fig. 2b) where, in the warmer doubled-CO₂

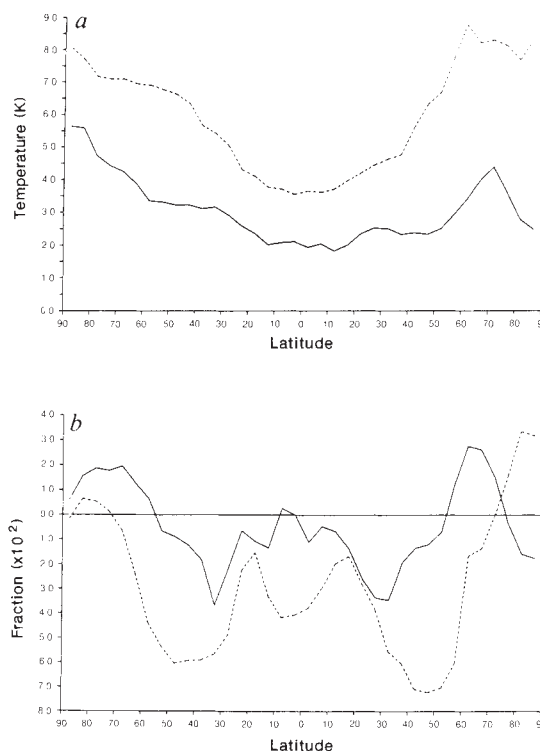


FIG. 1 Zone-averaged equilibrium changes in a, surface temperature and b, fractional cloud cover on doubling CO₂ (five-year means). Solid line, simulation with CW. Dashed line, simulation with RH.

simulation, ice cloud which is depleted rapidly by precipitation is replaced by water cloud which is depleted more slowly, so that cloud cover increases. This can be seen more clearly in the height-latitude distribution of the changes in cloud liquid water and cloud ice (Fig. 3). Also shown are the mean 0 °C and -15 °C isotherms in the control simulation. Because of the increases in cloud, the enhanced absorption of solar radiation in mid-latitudes found with RH is greatly reduced when using CW (Fig. 4).

One can define a climate sensitivity parameter λ ($=\Delta Q/\Delta T$) where ΔQ is the change in radiative heating at the tropopause on doubling CO_2 and ΔT is the equilibrium change in surface temperature. Assuming¹⁰ $\Delta Q = 4 \text{ W m}^{-2}$, λ is increased from 0.77 to $1.48 \text{ W m}^{-2} \text{ K}^{-1}$ on replacing RH by CW. Thus changes in the phase of cloud water provide a strong negative feedback on climate, cancelling almost all the previous strong positive feedback from cloud in RH¹¹. Although our model does not allow for this process in deep convective clouds, these have a small areal coverage and a relatively high liquid-water content,

however, suggesting that any radiative feedback from this type of cloud would be small.

The lower-cloud increases in the simulations arise from the assumption that under similar conditions ice is removed more efficiently from cloud than water. In real water clouds, relatively few droplets become large enough to fall out, as the available condensed water is spread over many cloud condensation nuclei (CCN)¹². Very few CCN act as ice nuclei¹³, so in ice clouds the cloud water is spread over fewer, larger particles, all of which have appreciable fall speeds¹⁴. The microphysics of mixed-phase clouds are complex and not fully understood. Although our model does not explicitly represent them it effectively includes their most important aspect, the continual rapid conversion of water to ice¹⁵. A more complex representation would certainly change the pattern in Fig. 2 locally, but the vertically integrated response would probably depend primarily on the overall change from pure ice cloud to pure water cloud. Thus our assumptions concerning the relative fall-out of ice and water are at least qualitatively realistic.

As 1 m s^{-1} may be a large value for ice fall speed, we have rerun experiment CW using a parameterization of fall speed in terms of ice-water content based on observations¹⁴ (experiment CWH). For typical simulated ice-water contents, this gives fall speeds of 0.2 m s^{-1} for high ice cloud and 0.7 m s^{-1} for ice cloud near the freezing level. Cloud changes (Fig. 2c) are qualitatively similar to Fig. 2b and the global warming is 3.2 K, only slightly larger than in CW. The similarity between CW and CWH is to be expected, given that the depletion rate of ice relative to water is large near the freezing level in both experiments.

It has recently been suggested that changes in cloud radiative properties may contribute an important feedback to climate^{4,5,16}. In a warmer climate, greater cloud water content would increase cloud albedos (a negative feedback) and emittances (a positive feedback). A further experiment was carried out using the revised version of the model with an additional modification to

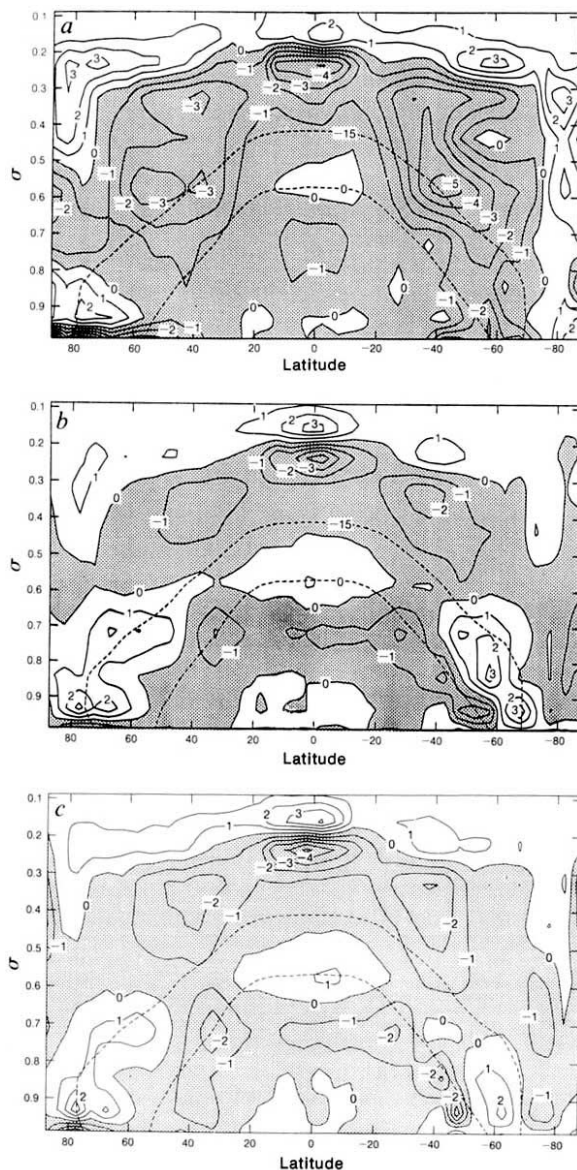


FIG. 2 Height-latitude cross-section of equilibrium changes in cloud amount on doubling CO_2 . The height coordinate is σ (pressure/surface pressure). Areas of reduction are stippled and the contours are every 1%. The dashed lines are the 0 and -15 °C contours from the control simulation. *a*, Simulation with RH. *b*, Simulation with CW. *c*, Simulation with CWH.

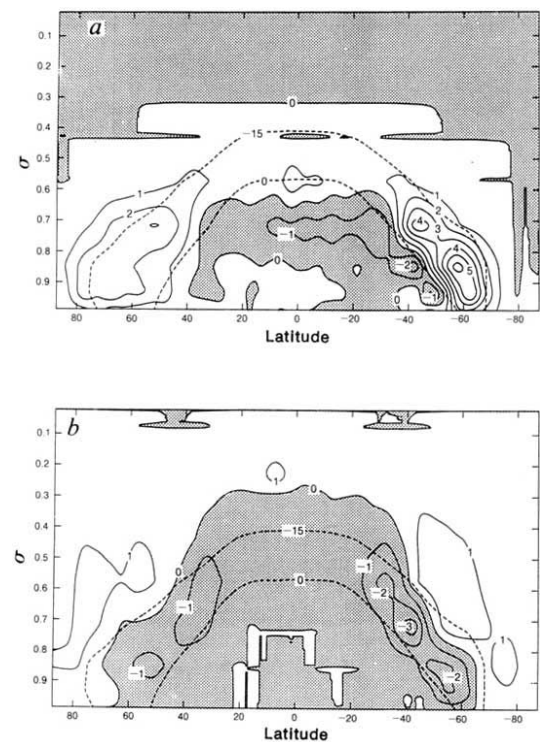


FIG. 3 Height-latitude cross-section of equilibrium changes in cloud water in CW on doubling CO_2 . The height coordinate is σ . Areas of reduction are stippled. The dashed lines are the 0 and -15 °C contours from the control simulation. *a*, Liquid-water content only. Contours every $4 \times 10^{-6} \text{ kg kg}^{-1}$. *b*, Ice-water content only. Contours every $4 \times 10^{-7} \text{ kg kg}^{-1}$.

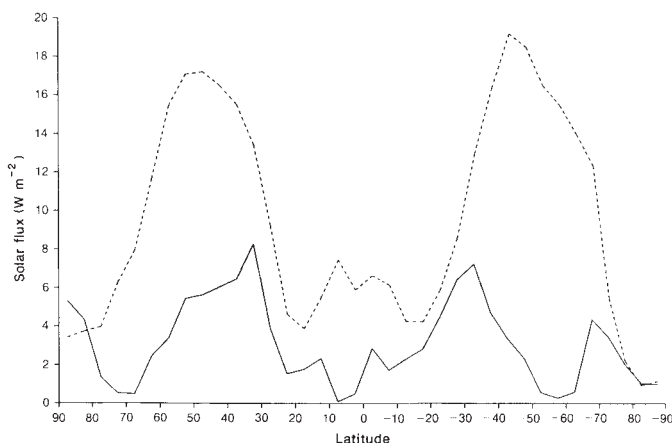


FIG. 4 Zone-averaged equilibrium change in net solar flux at the top of the atmosphere on doubling CO_2 (five-year mean). Solid line, simulation with CW. Dashed line, simulation with RH.

allow the cloud radiative properties to depend on the cloud water content (scheme CWRP). (Given the similarity between CW and CWH, this experiment was carried out with the CW formulation only.) Analytical fits of the short-wave albedo and transmittance as functions of cloud-water path and solar zenith angle were made to a published scheme¹⁷ based on observations and theoretical considerations. The long-wave flux emittance was given by $(1 - \exp(-\kappa q))$ where q is the condensed-water path and $\kappa = 0.13 \text{ m}^2 \text{ g}^{-1}$ for water clouds and $0.065 \text{ m}^2 \text{ g}^{-1}$ for ice clouds¹⁸.

On doubling the CO_2 concentrations the global annual average equilibrium surface temperature increased by 1.9 K with CWRP compared to 2.7 K with prescribed cloud radiative properties (CW), indicating a further increase in λ to $2.1 \text{ W m}^{-2} \text{ K}^{-1}$. This negative feedback from changes in cloud radiative properties lies between the very strong negative feedback suggested by one-dimensional calculations^{16,19} and the positive feedback found in a previous study with a low-resolution general-circulation model (GCM)⁴⁻⁶. We note that all existing numerical studies assume that the number of CCN remains constant under climate change. Changes in CCN are a further possible source of feedback^{7,20}.

The sensitivity of the climate to doubling the CO_2 concentrations that we have found is substantially smaller than the 4–5 K found in other recent studies^{10,21}. A significant part of this reduction in sensitivity is due to the different cloud scheme. If correct this will have two important consequences. First, the equilibrium warming expected from the estimated 2 W m^{-2} increase in radiative heating at the tropopause since 1860 (ref. 9) will be smaller. Second, the proportion of the equilibrium reached will be greater, because the thermal response time of the climate system is reduced if climate sensitivity is reduced²². Based on calculations using a simple one-dimensional model of the ocean, the expected warming to date, using the climate sensitivity found when employing CW, is 0.6 K. This is close to the observed warming since the beginning of the century (0.5 K; ref. 23). The record of observations shows considerable variability on interannual and interdecadal timescales, however, including a period of cooling during the 1950s and 1960s, so that it is not possible to attribute the 0.5 K rise unambiguously to the effect of increases in trace gases. □

5. Roeckner, E. *Nature* **335**, 304 (1988).
6. Schlesinger, M. E. *Nature* **335**, 303–304 (1988).
7. Charlson, R. J. *et al.* *Nature* **334**, 418–420 (1988).
8. Mitchell, J. F. B. & Warrilow, D. A. *Nature* **330**, 238–240 (1987).
9. Smith, R. N. B. *Q. J. R. Met. Soc.* (in the press).
10. Dickinson, R. E. & Cicerone, R. J. *Nature* **319**, 109–115 (1985).
11. Wilson, C. A. & Mitchell, J. F. B. *J. geophys. Res.* **92**, 13,315–13,343 (1987).
12. Pruppacher, H. R. & Klett, J. D. *Microphysics of Cloud and Precipitation*. (Reidel, Dordrecht, 1978).
13. Heymsfield, A. J. & Sabin, R. M. *J. Atmos. Sci.* **46**, 2252–2264 (1989).
14. Heymsfield, A. J. *J. Atmos. Sci.* **34**, 367–381 (1977).
15. Mason, B. J. *The Physics of Clouds* (Clarendon, Oxford, 1957).
16. Somerville, R. C. J. & Remer, L. A. *J. geophys. Res.* **89**, 9668–9672 (1984).
17. Liou, K.-N. & Wittman, G. D. *J. Atmos. Sci.* **36**, 1261–1273 (1979).
18. Stephens, G. L. *J. Atmos. Sci.* **35**, 2123–2132 (1978).
19. Somerville, R. C. J. & Iacobellis, S. *Am. Met. Soc. 7th Conf. Ocean–Atmos. Interaction* 53–55 (1988).
20. Liou, K.-N. & Ou S.-C. *Preprints of 10th Int. Cloud Phys. Conf.* Vol. II, 118–120 (1988).
21. Schlesinger, M. E. & Mitchell, J. F. B. *Rev. Geophys.* **25**, 760–798 (1987).
22. Hansen, J. E. *et al.* *Geophysical Monograph Series* Vol. 29, 130–163 (American Geophysical Union, Washington, DC, 1984).
23. Jones, P. D. *et al.* *Nature* **338**, 790 (1988).

ACKNOWLEDGEMENTS. We would like to thank R. Smith who devised and wrote the cloud water scheme used.

Disequilibrium silicate mineral textures: fractal and non-fractal features

Anthony D. Fowler*, H. Eugene Stanley† & Gérard Daccord‡

* Ottawa–Carleton Geoscience Centre, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5

† Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

‡ Dowell Schlumberger, Z.I. Molina, BP 90, 42003 St Etienne, Cedex 1, France

IGNEOUS rocks formed from lava flows of the Archaean era (>2,700 million years ago) are often found to contain disequilibrium-textured crystals characterized by spherulitic, branching or dendritic morphologies that occur in layers near the flow surface. Well-known examples are the plagioclase spherulites of basalts and the platy and branching spinifex-textured olivines and pyroxenes of komatiites^{1,2}. Here we present evidence that, over a finite range of length scales, some disequilibrium textures are scale invariant. This observation implies that over this range of length scales their random patterns can be quantitatively characterized by a unique number, the fractal dimension³. We also demonstrate that some textures have a crossover from fractal to non-fractal behaviour. It is known that most disequilibrium crystals arise in part from rapid cooling and represent the case where the growth rates of the crystals are large compared to the diffusion rates in the silicate melt^{1,2,4}. We therefore formulate a quantitative model for the growth that is based on a variant of diffusion-limited aggregation (DLA)⁵.

Figure 1 shows a section of small disequilibrium crystals, at different magnifications. These form close to the quench/glassy margin of some lava flows⁶, and their outline remains the same as the microscope objective power is increased. Similar remarks can be made about textures on other length scales, such as the much larger branching pyroxene crystal of Fig. 2a.

The concepts and techniques of fractal geometry³ have proven useful in understanding other random morphologies, in part because a fractal object is characterized quantitatively by a number, the fractal dimension d_f , that relates the increase in mass M of the object to its characteristic length scale L , $M \approx L^{d_f}$ (ref. 7).

To determine d_f we digitized photographs of the textures using a video camera with a grid of 65,536 pixels. First we randomly chose any pixel belonging to the digitized texture. About this 'local origin' we constructed a sequence of concentric shells of

Received 14 June; accepted 14 August 1989.

1. Manabe, S. & Wetherald, R. T. *J. Atmos. Sci.* **24**, 241–259 (1967).
2. Schneider, S. H. *J. Atmos. Sci.* **29**, 1413–1422 (1972).
3. Wetherald, R. T. & Manabe, S. *J. Atmos. Sci.* **45**, 1397–1415 (1988).
4. Roeckner, E. *et al.* *Nature* **329**, 138–140 (1987).

radius r . For each shell, we formed the correlation function $\mathcal{C}(r)$ by dividing the number of pixels belonging to the texture by the total number of pixels in that shell. To obtain good statistics we repeated the calculation using all texture pixels as local origins.

The correlation function $\mathcal{C}(r)$ gives the probability that a pixel separated from the local origin by a distance r is part of the pattern. We expect this function to scale in the same way as the density of pixels in a two-dimensional fractal object⁷.

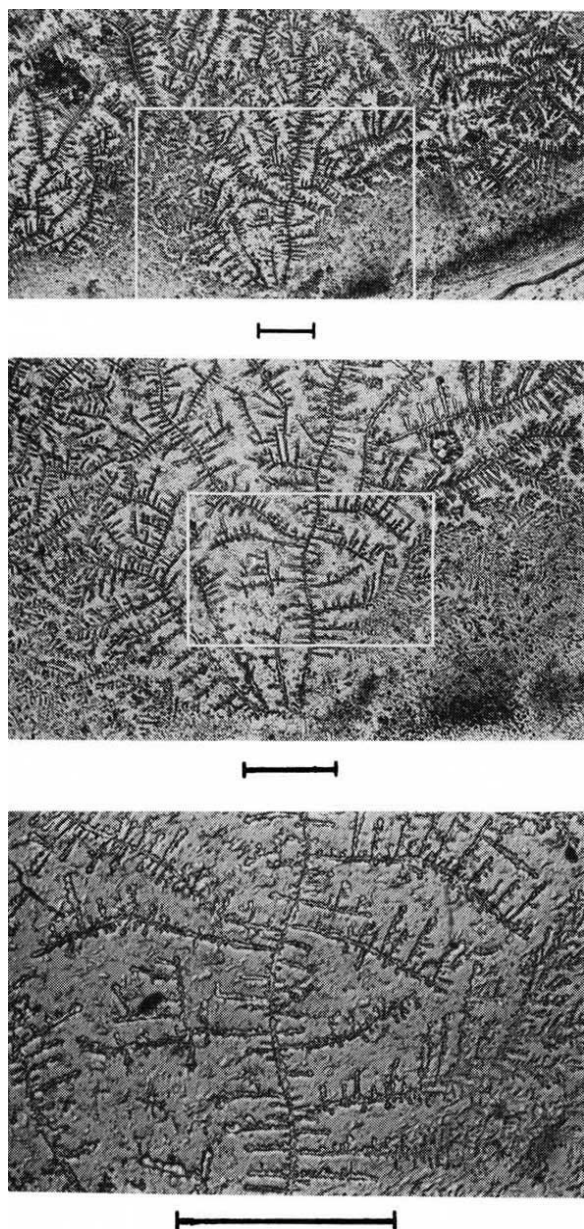


FIG. 1 Photomicrographs taken at increasing magnifications of the same field, of branching crystals (interpreted to be olivine) sampled from beneath the quenched pillow margin of an Archaean metabasalt flow. The scale bars represent a length of 0.1 mm. The enclosed areas indicate the field of view of the subsequent image. The branching pattern is independent of magnification; this scale invariance is characteristic of fractal objects. The images were taken using Nomarski interference microscopy of acid-etched samples¹⁵. Pillows may form during subaqueous eruption when lava in contact with water quickly quenches to a glass. The glass remains plastic and deforms under pressure of the flow into a bulbous shape (pillow) in a manner analogous to blown glass, except that the interior is lava filled. The pillows often contain a spectrum of mineral growth morphologies from the glassy margin to the slowly cooled pillow centre (~ 0.5 m).

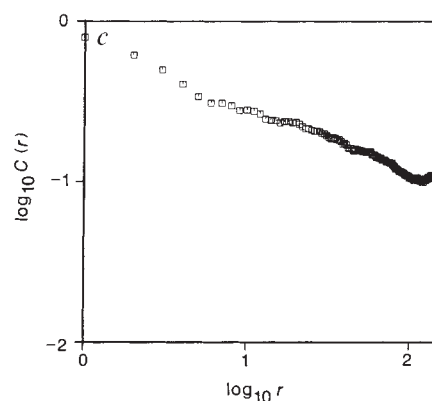
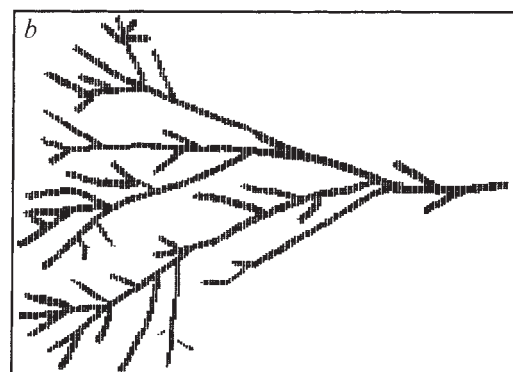


FIG. 2 *a*, Photograph of a large (scale = 32 cm) branching pyroxene from a gabbro sill in Munro Township of the Archaean Abitibi Subprovince, Ontario. These branching crystals are most evident on weathered surfaces and are exposed over several hundred metres of the sill in a haphazard manner. Elsewhere, such as at the location of the photo, their long axes are normal to cooling contacts. The crystal is clearly scale invariant and fractal in form. *b*, is the digitized form and *c*, shows that $d_f = 1.60 \pm 0.10$.

That is, the number of pixels that form part of the object is not proportional to the square of r but to the power d_f ,

$$\mathcal{C}(r) \approx r^{d_f-2} \quad (1)$$

Thus the slope of a double logarithmic plot of $\mathcal{C}(r)$ against r gives a quantitative value of d_f .

Figure 3a shows a digitized version of the olivine of Fig. 1. Figure 3b is a double logarithmic plot of $\mathcal{C}(r)$ against r , and shows that for r less than a crossover value r_c of about 70 pixels, the data fall on a line of slope -0.20 , corresponding to $d_f = 1.80 \pm 0.05$. For $r > r_c$, $\mathcal{C}(r)$ has a constant 'saturation' value $\mathcal{C}_{\text{sat}}$; hence over this range the object is non-fractal. We can check the numerical value of $\mathcal{C}_{\text{sat}}$ by calculating the mean density of the pattern; we find 6,682 black pixels in a total area of $222 \times 201 \approx 44,000$. Hence we expect $\log \mathcal{C}_{\text{sat}} \approx \log(0.15) \approx -0.82$ indicating that the initial concentration of olivine 'components' in the liquid was on the order of 15% in accordance with measured estimates of the sum of the contents of FeO and MgO of the rock⁶.

We analysed at least a dozen additional crystals and consistently found evidence for their fractal nature below a crossover radius r_c . For example, the pyroxene (Fig. 2) has $d_f = 1.60 \pm 0.10$. The larger statistical uncertainty is due to the fact that outcrop weathering has somewhat obscured the outline of the crystal.

Experimental studies have clearly documented the conditions responsible for the growth of disequilibrium silicate crystals, but an understanding of the physical mechanisms remains obscure. We note that the fractal dimensions we find are similar to those found for DLA. The DLA algorithm starts with a seed particle and releases a random walker from the periphery of a circle enclosing the seed. When the random walker touches the seed, it sticks and a second walker is released. This process is

iterated until a large aggregate is formed. The random walkers in DLA correspond to structural units in the silicate melt which undergo Brownian motion and upon collision attach to a central nucleus without rearrangement. The model is thus somewhat analogous to crystal growth in a magma under conditions of high supercooling. Protuberances preferentially grow because the probability is small that a molecule in a highly supercooled liquid will travel without collision into the 'fjords' between the branches of the crystals.

These disequilibrium crystals share other characteristics with DLA in addition to having a similar value of the fractal dimension. For example, they display a ramified (branched) structure, starting from a unique seed. Moreover, the branches do not form any 'closed loops'. A main difference between the DLA model and actual physical agglomeration processes is that the concentration of particles tends to zero in the DLA model. In any physical system, however, there is a finite concentration of molecules. This leads to the possibility that growth can occur in several different places at the same time. The effect⁸ of deviation from this limiting condition of zero concentration on the structure of the final patterns is the crossover from fractal behaviour at small length scales to non-fractal (that is, constant density) behaviour at larger length scales. The characteristic length at which this crossover occurs depends on the concentration of particles and seeds.

Figure 4 shows the results of a 'finite-concentration' simulation, with the added feature that anisotropy has been introduced. Several seed particles are introduced and a non-zero concentration of walkers is distributed at random. At each time step, each walker attempts to move to one of its four neighbouring sites (chosen at random). If the neighbouring site is empty, the move is allowed, otherwise the move is prohibited. When a walker steps on a perimeter site of any cluster, it sticks irreversibly.

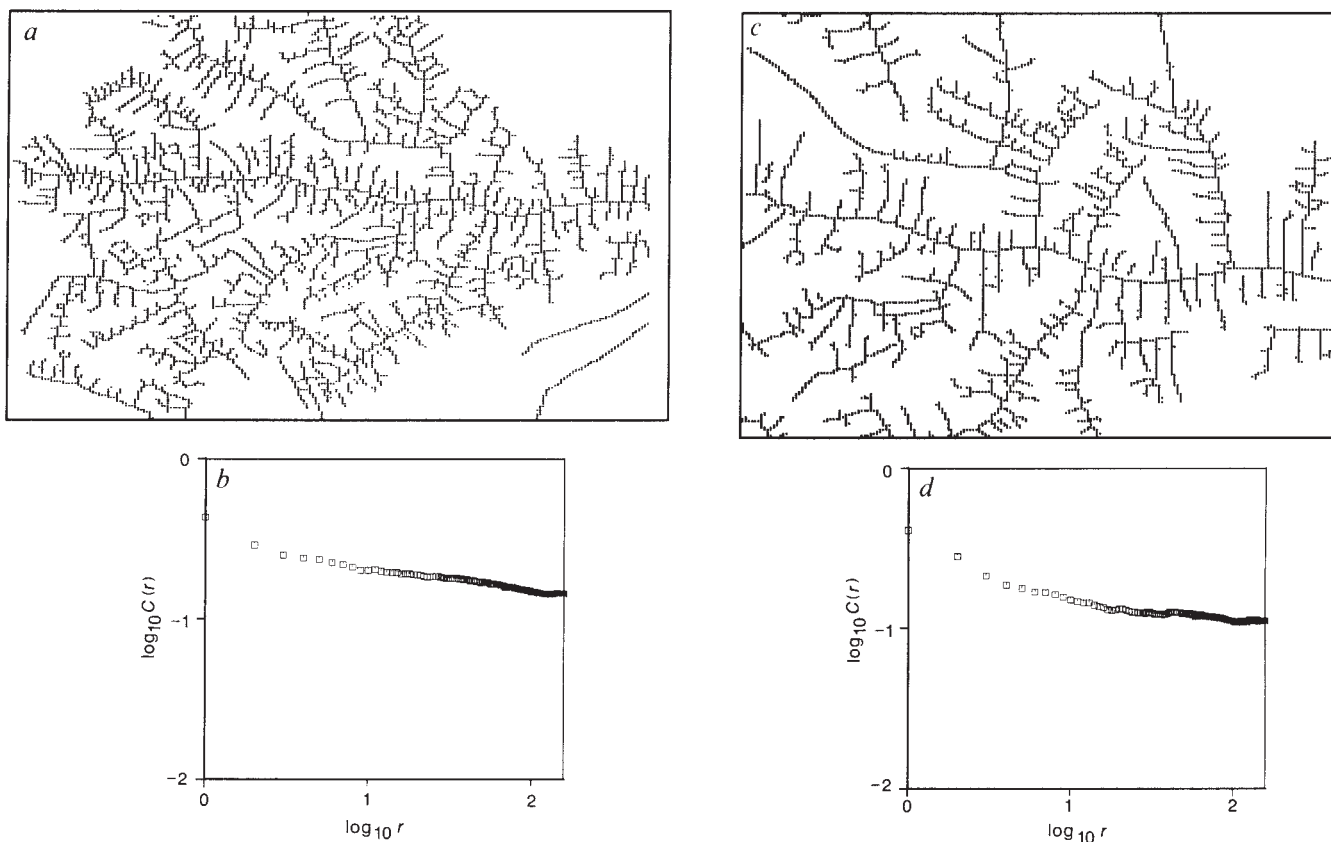


FIG. 3 *a*, Digitized representation of the central area of Fig. 1*a*. *b*, The graph of the correlation function as a function of length is characteristic of fractal forms and yields $d_f = 1.80 \pm 0.05$. *c* and *d*, The digitized forms of the central

area of Fig. 1*b* and the graph of correlation function as a function of length respectively, $d_f = 1.72 \pm 0.07$. Note that the fractal dimensions d_f of the two images are almost identical under change of scale.

Anisotropy is present in disequilibrium-crystal growth. The anisotropy is assumed to start when the fastest growing face of the crystal nucleus is aligned with the thermal or compositional gradient. Once initiated, the preferred direction is conserved as the crystals grow out from the depleted area in the direction of highest concentration, normal to the cooling contact. We have incorporated anisotropy in our model by introducing a local noise reduction. In conventional global noise reduction, a particle sticks after visiting a perimeter site s times (see, for example, ref. 9). In contrast, for local noise reduction, the parameter s has a lower value for perimeter sites co-linear with cluster sites; for the simulation of Fig. 4, we chose $s = 1$ for perimeter sites aligned with the previous direction of growth and $s = 2$ for the remaining perimeter sites.

The similarity of the olivine photographs (Fig. 1) and the simulation (Fig. 4) is rather striking, both qualitatively and quantitatively (they have roughly the same value of d_f , followed at a larger distance by a crossover to constant mass density). At small radii the short distance between tips of the mineral relative to the diffusion length of the molecules prevents in-filling, whereas at larger radii in the presence of a finite number of molecules the diffusion length is short with respect to tip distance so in-filling growth becomes more probable.

Although the stochastic model presented, DLA, provides us with a physical mechanism of growth, it yields no quantitative information about the actual mineral growth. A kinetic model that is a variant of DLA may be used, however, to calculate parameters associated with the growth of the olivine at the

crossover. We use the Smoluchowski formulation, a theory of anomalous reaction kinetics that accounts for diffusion-limited growth. We base our analysis on the finding⁸ that this crossover is entirely due to the finite concentration of particles. When the crossover occurs, the average displacement of a particle during a period of time δt equal to the mean time between two particles sticking on the cluster is of the order of the size of the cluster,

$$\sqrt{D \delta t} = r_c \quad (2)$$

To calculate δt we assume a given diffusivity D for the particles and thereby obtain an estimate of the particle concentration and other physical parameters.

We assume the flux of particles sticking on the cluster is the same as if the cluster were a filled sphere of size R ; this assumption has been verified to hold for DLA¹⁰, provided R is equal to an effective radius of the DLA cluster. This effective radius is smaller than the DLA actual outer radius but of the same order of magnitude. Therefore we use the expression of the flux of incoming particles toward a spherical sink^{11,12}

$$\phi = 4\pi D R C \left(1 + \frac{R}{\sqrt{\pi D t}} \right) \quad (3)$$

where ϕ is the flux (number of particles per unit of time), R the cluster radius (effective radius of the DLA cluster), C the particle concentration far from the cluster and t the time since the beginning of the growth.

For DLA simulations the mean time between growth events δt corresponds to an increase in branch length equal to the particle size. The mean branch width for DLA is almost equal to the particle size¹³. Although the branch width W of real physical objects such as the minerals discussed here remains essentially constant, W^3 is much larger than the size of an individual molecule. Hence the corresponding mean time δt required to increase the branch length by W is simply $1/\phi$ multiplied by the number of molecules A , in a cube of volume W^3 of the mineral. Thus

$$\delta t = \frac{A}{\phi} \quad (4a)$$

and

$$A = W^3 (\rho/M) N_A \quad (4b)$$

where ρ and M are the mineral density and relative molecular mass respectively, and N_A is Avogadro's number. For long times (for example, the time t required to reach $R = r_c$), equation (3) reduces to $\phi = 4\pi D r_c C$ and from equation (4a)

$$\delta t \approx \frac{A}{4\pi D r_c C} \quad (5)$$

The concentration of species at the crossover may be calculated because $\delta t = r_c^2/D$ from equation (2). Thus equation (5) becomes

$$C \approx \frac{A}{4\pi r_c^3} \quad (6)$$

From Fig. 3b we see that $r_c = 67$ pixels, which corresponds to 0.022 cm in Fig. 3a. Because a branch width is one pixel, $W = r_c/67 = 3.3 \times 10^{-4}$ cm. Using equation (4b), we obtain $A = 4.25 \times 10^{11}$ molecules. Hence from equation (6), $C = 3 \times 10^{15}$ molecules cm^{-3} . Using $D = 5 \times 10^{-7} \text{ cm}^2 \text{ sec}^{-1}$, we calculate from equation (5) that the time δt of growth at the crossover is 970 s, yielding an instantaneous growth rate of $3.4 \times 10^{-7} \text{ cm s}^{-1}$, similar to that reported¹⁴ for skeletal olivine growth. Because the growth rate of DLA slows with growth, our value is a lower bound.

These results should be considered only as order of magnitude estimates because we have used very rough assumptions, the most critical of which is the use of equation (3). In the absence of any quantitative theory for DLA, however, we could not find an alternative approach. $\square$

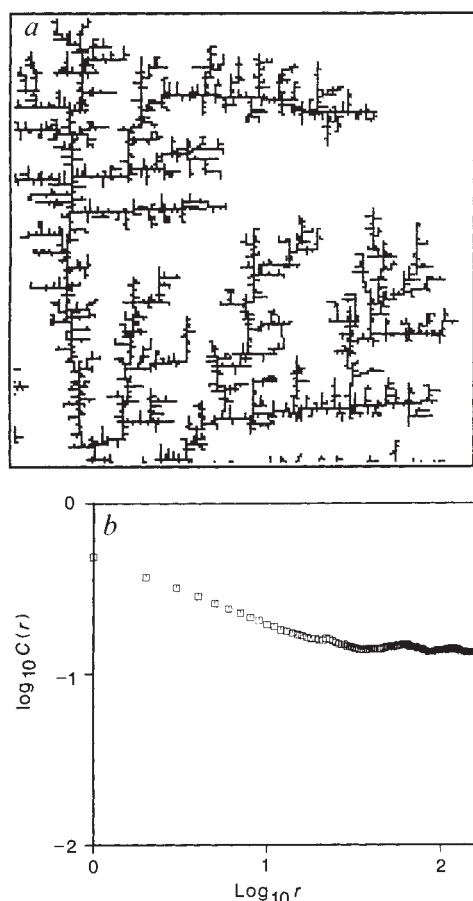


FIG. 4 *a*, Pattern simulated using the algorithm of ref. 8 modified to include local noise reduction ($s = 2$ here). Of a total of 32,400 available pixels, 4,512 were selected as mobile particles. Only the central part of the simulation is illustrated because of boundary effects. *b*, The plot of the density correlation function, showing that in the fractal region $d_f = 1.73$. The crossover to non-fractal behaviour occurs at $r \approx 30$.

Received 25 August 1988; accepted 7 August 1989.

1. Lofgren, G. *Am. J. Sci.* **274**, 243–273 (1974).
2. Donaldson, C. H. *Contr. Miner. Petrol.* **57**, 187–213 (1976).
3. Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, New York, 1983).
4. Keith, H. D. & Padden, F. J. Jr *J. appl. Phys.* **34**, 2409–2421 (1963).
5. Witten, T. A. & Sander, L. M. *Phys. Rev. Lett.* **47**, 1400–1403 (1981).
6. Fowler, A. D., Jensen, L. S. & Pelouquin, A. S. *Can. Mineralogist* **25**, 275–289 (1987).
7. Stanley, H. W. *Fractals and Multifractals: An Introduction* (Oxford University Press, in the press).
8. Witten, T. A. & Meakin, P. *Phys. Rev.* **B28**, 5632–5642 (1983).
9. Nittmann, J. & Stanley, H. E. *Nature* **321**, 663–668 (1986).
10. Yang, C.-L., Chen, Z.-Y. & El-Sayed, M. A. *J. phys. Chem.* **91**, 3002–3006 (1987).
11. Collins, F. C. & Kimball, G. E. *J. Colloid Sci.* **4**, 125–137 (1949).
12. Rice, S. A. *Diffusion-Limited Reactions* (Elsevier, New York, 1985).
13. Daccord, G. *Phys. Rev.* **A39**, 1365–1368 (1989).
14. Donaldson, C. H. *Lithos* **8**, 163–174 (1975).
15. Clark, A. H., Pearce, T. H., Roer, P. L. & Wolfson, I. *Am. Miner.* **71**, 734–741 (1986).

ACKNOWLEDGEMENTS. We thank R. Ernst, R. Kretz, G. Lofgren & J. Veizer for providing constructive criticism and T. H. Pearce for access to the Nomarski microscope at Queen's University. We acknowledge the financial support of the University of Ottawa Development Fund, the Natural Science and Engineering Research Council of Canada, the NSF and the Office of Naval Research.

Carbon isotopes in soils and palaeosols as ecology and palaeoecology indicators

T. E. Cerling, J. Quade, Y. Wang & J. R. Bowman

Department of Geology and Geophysics, University of Utah, Salt Lake City, Utah 84112, USA

THERE are few quantitative techniques in use today for palaeoecological reconstruction in terrestrial depositional systems. One approach to such reconstructions is to estimate the proportion of C_3 to C_4 plants once present at a site using carbon isotopes from palaeosol carbonates^{1–3}. Until now, this has been hampered by an inadequate understanding of the relationship between the carbon isotopic composition of modern soil carbonate and coexisting organic matter. Here we have found that the two systematically differ by 14–16‰ in undisturbed modern soils. This difference is compatible with isotopic equilibrium between gaseous CO_2 , and aqueous and solid carbonate species in a soil system controlled by diffusive mass transfer of soil CO_2 derived from irreversible oxidation of soil organic matter. Organic matter and pedogenic carbonate from palaeosols of Pleistocene to late Miocene age in Pakistan also differ by 14–16‰. This indicates that diagenesis has not altered the original isotopic composition of either phase, thus confirming their use in palaeoecological reconstruction.

Stable carbon isotopes in pedogenic carbonate^{1–4} and soil organic matter^{5,6} can be used as palaeoecological indicators. A critical assumption, however, in the use of pedogenic carbonate is that of isotopic equilibrium between the carbon species—soil CO_2 , dissolved carbonate species, and solid carbonate—in the soil system buffered by an effectively infinite soil CO_2 reservoir. Other studies in recent years have invoked processes such as 'Rayleigh' distillation^{7,8}, kinetic fractionation⁸, and partial exchange or inheritance^{8,9} to describe carbon isotope systematics in soils. The principal drawback of the latter models is that little isotope data on co-genetic organic carbon and soil carbonate was collected. Instead, the isotopic composition of the organic matter was usually assumed^{7–9}. Where measured, the results were considered unreliable because of the vegetation changes resulting from cultivation⁸. Many of the studied soils have complex vegetational histories attributable to either the effects of cultivation and extensive grazing^{7–9} or, in the case of calcretes⁷, to long periods of development in the presence of pre-Holocene vegetation. These complexities make interpretation of the soil isotope system exceedingly difficult.

Our approach is to study ten soils in North America with simple pedogenic histories (Table 1). All the soils are latest Pleistocene or, in most cases, entirely Holocene in age. In all

but one case, vegetation cover can be assumed to have been constant and similar to that of today. The six mid-western soils are located on national prairie preserves where the vegetation shows little or no post-settlement impact. The Nevada soils come from remote mountain settings that are unsuitable for grazing, in which vegetation remains entirely native. Vegetation at the New York site, though disturbed, must always have been 100% C_3 in character given the forested nature of the region.

The isotopic composition of plants using the C_3 and C_4 photosynthetic pathways have $\delta^{13}C$ values averaging ~ -27 and -13% , respectively¹⁰. We have examined ten soils the biomasses of which span the range from nearly 100% C_3 to 100% C_4 and have analysed coexisting bulk-soil organic matter and pedogenic carbonate in each. The soil carbonates were present as nodules and as pendants on pebble undersides. We confined our studies to areas where the rainfall exceeds 35 cm yr^{-1} because modelling of soils shows that pedogenic carbonate from soils with very low respiration rates, such as in deserts, can be isotopically enriched because of diffusional mixing with the atmosphere¹¹. We sampled carbonates at depths $>30\text{ cm}$ because the near-surface carbonate could be affected by diffusional mixing with atmospheric CO_2 , irrespective of the soil respiration rate. Nearly all the soils studied contained carbonate in the parent alluvium, loess or till.

The diffusion coefficients for $^{12}CO_2$ and $^{13}CO_2$ differ, producing a 4.4‰ enrichment in the plant-derived component of soil CO_2 compared with respired CO_2 (refs 1, 11, 12). At equilibrium, the isotopic equilibrium factor $10^3 \ln \alpha_{CO_2\text{-calcite}}$ is -9.8 to -12.4 for the temperature range $25-0^\circ\text{C}$ (ref. 13). Pedogenic carbonate precipitated in equilibrium with soil CO_2 should, therefore, be $\sim 14\%$ (25°C)– 17% (0°C) enriched relative to respired CO_2 . At low soil respiration rates, carbonates precipitated in equilibrium with soil CO_2 would have an even greater enrichment as a result of the atmospheric effects mentioned above.

The isotopic composition of pedogenic carbonate from a single soil in the Konza Prairie of Kansas is constant in the soil from depths of 40 to 103 cm, and it is 14.5‰ more positive than the co-existing organic material in the soil (Fig. 1). This difference supports the model of a diffusion-controlled soil CO_2 - $CaCO_3$ system in isotopic equilibrium. The patterns displayed by this soil and those described below are not consistent with kinetic or Rayleigh distillation processes resulting in isotopic disequilibrium between soil CO_2 and precipitated carbonate. Similarly, modern soil organic matter and coexisting pedogenic carbonate from nine other North American soils differ by $\sim 14-16\%$ (Fig. 2), as predicted for high-respiration-rate soils. All soils we studied, including palaeosols from Pakistan, show strong leaching and contain well developed organic horizons, thus indicating high respiration rates^{14,15}. In low-respiration-rate soils, such as in deserts, diffusional mixing of atmospheric CO_2 and plant-derived CO_2 can produce up to 2‰ enrichment in the $\delta^{13}C$ of soil carbonate at 50-cm depth¹¹. No systematic trend in

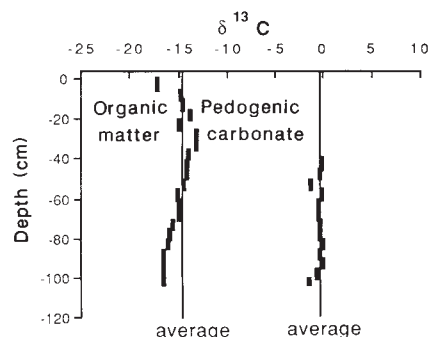


FIG. 1 The isotopic composition of soil organic matter and pedogenic carbonate as a function of depth in a prairie soil from Kansas, USA.

TABLE 1 General features of soil sites

Locality	Soil series	Soil type	$\delta^{13}\text{C}$ (PDB*) org. matter	$\delta^{13}\text{C}$ (PDB*) carbonate	MAP†	MAT‡	Carbonate sample depth (cm)	Parent material	Vegetation
Manhattan, Kansas	Dwight	Natrustoll	-15.7§	-1.6 ± 0.4	81	13.3	50-86	loess	native prairie
Manhattan, Kansas	Tully	Argiustoll	-14.9 ± 1.1	-0.4 ± 0.3	81	13.3	40-103	colluvium	native prairie
Lincoln, Nebraska	Burchard	Argiudoll	-17.3§	-2.4 ± 0.7	68	11.4	94-129	till	native prairie
Pocahontas, Iowa	Clarion	Hapludoll	-18.5§	-3.4 ± 0.9	73	8.7	106-171	till	native prairie
Watertown, South Dakota	Barnes	Haploboroll	-17.3 ± 1.1	-2.0 ± 0.4	53	6.2	35-65	till	native prairie
Saskatoon, Saskatchewan, Canada	Elstow	Haploboroll	-24.2 ± 0.7	-7.9 ± 0.6	35	1.6	40-75	lake deposits	native prairie
Binghamton, New York	Howard	Hapludalf	-25.6¶	-9.4 ± 0.2	92	7.7	100-125	outwash gravel	mixed woodland
Spring Mountains-4, Nevada	—	Calciorthid	-23.4#	-6.8 ± 1.1	55	8.8	40-60	alluvial gravel	pigmy conifer woodland
Spring Mountains-5, Nevada	—	Calciorthid	-23.7#	-8.6 ± 0.2	61	7.3	40-60	alluvial gravel	conifer woodland
Spring Mountains-6, Nevada	—	Calciorthid	-23.9#	-8.5 ± 0.5	68	5.7	40-60	alluvial gravel	conifer woodland

* PDB=the Pee Dee-belemnite standard CO_2 for carbon and oxygen isotopic composition.

† MAP=mean annual precipitation (cm).

‡ MAT=mean annual temperature ($^{\circ}\text{C}$).

§ ~5 g soil from 5-10 cm intervals over whole profile except surface litter, mixed together, analysis on a single homogenized sample.

|| ~5 g soil from 5-10 cm intervals over whole profile except surface litter, analysed separately, results averaged.

¶ Soil sample from 15-35 cm only, homogenized before analysis.

Soil sample from 40-60 cm only, homogenized before analysis.

Fig. 2 would be expected for soils in which the vegetation has been altered by agricultural or pastoral practices or, in the case of old surface soils, by major climatic change. The isotopic compositions of all the soils studied are compatible with a diffusion-controlled system in which all oxidized carbon species are in isotopic equilibrium. Therefore, the isotopic composition of soil organic matter or pedogenic carbonate in high-respiration-rate soils is a direct indicator of the fraction of the biomass using the C_3 or C_4 photosynthetic pathways.

Obtaining palaeoecological information from palaeosol carbonates and organic matter requires that the sampled palaeosols bear the imprint of time spans during which vegetation remained undisturbed by humans or by major climate change, as in the case of the modern soils we selected for study. Clearly, buried pre-Holocene palaeosols of any type should not have experienced anthropogenic modification. We estimate that the length of pedogenesis for Pliocene and Pleistocene soils in Pakistan to have been <10,000-15,000 yr, and probably much less in most cases. Thus, the carbonate build-ups in these palaeosols should not be, in most instances, polygenetic. Palaeosol humus probably represents organic matter from the last few hundred years before

burial, given the short residence times typical for humus in most modern soils¹⁵.

Figure 2 also shows the isotopic composition of co-existing palaeosol organic matter and pedogenic carbonates from palaeosols in the Siwalik sequence of Pakistan. The concordance with the modern trend indicates that the pedogenic carbonates in the palaeosols have not been isotopically altered during diagenesis, notwithstanding several kilometres of burial. The isotope results from two Pakistani palaeosols indicate the highest soil temperatures of any soil or palaeosol studied. Mean annual temperatures and maximum soil temperatures (at 50-cm depth) for the Potwar Plateau of Pakistan are ~24 and 33 $^{\circ}\text{C}$, respectively, much higher than for any North American site. Although this is interesting, we feel that use of co-existing carbonate-organic matter in palaeosols as a palaeotemperature indicator is premature, given the small sample size. Further evidence against diagenetic alteration is that modern soils and Miocene-to-Pleistocene palaeosol carbonates both range from -11 to +2‰, whereas coexisting organic matter for both range from ~-25 to -13‰.

Carbon isotope analysis of coexisting organic matter and pedogenic carbonate constitutes a simple test of the state of preservation of the original palaeoecological signal in soils. □

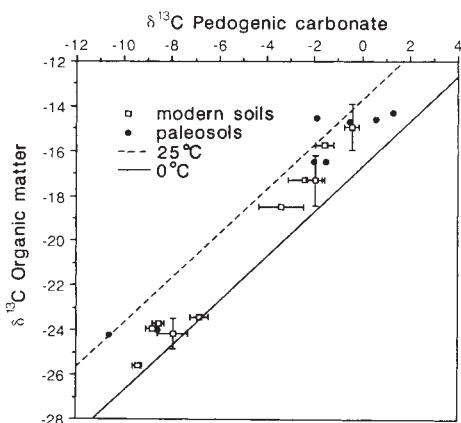


FIG. 2 The isotopic composition, relative to the PDB standard, of co-existing organic matter and pedogenic carbonate from ten modern North American soils, and from eight late Miocene-to-Pleistocene palaeosols in Pakistan. Solid lines represent isotope values for pedogenic carbonates in isotopic equilibrium with soil CO_2 derived from irreversible oxidation of soil organic matter in a diffusion-controlled soil system^{1,11} at 0 and 25 $^{\circ}\text{C}$.

Received 17 January; accepted 10 July 1989.

1. Cerling, T. E. *Earth planet. Sci. Lett.* **71**, 229-240 (1984).
2. Cerling, T. E. & Hay, R. L. *Quat. Res.* **25**, 63-78 (1986).
3. Cerling, T. E., Bowman, J. R. & O'Neil, J. R. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **63**, 335-356 (1988).
4. Amundson, R. G., Chadwick, O. A., Sowers, J. M. & Doner, H. E. *Soil Sci. Soc. Am. J.* **53**, 201-210 (1989).
5. Krishnamurthy, R. V., DeNiro, M. J. & Pant, R. K. *Nature* **298**, 640-641 (1982).
6. Krishnamurthy, R. V., Bhattacharya, S. K. & Kusumgar, S. *Nature* **323**, 150-152 (1986).
7. Salomans, W., Goudie, A. & Mook, W. G. *Earth Surf. Process.* **3**, 43-57 (1978).
8. Dever, L., Fontes, J. C. & Riche, G. *Chem. Geol.* **66**, 307-314 (1987).
9. Magaritz, M. & Amiel, A. J. *Soil Sci. Soc. Am. J.* **44**, 1059-1062 (1980).
10. Deines, P. in *Handbook of Environmental Isotope Geochemistry, 1. The Terrestrial Environment* 329-406 (Elsevier, Amsterdam, 1980).
11. Quade, J., Cerling, T. E. & Bowman, J. R. *Geol. Soc. Am. Bull.* **101**, 464-475 (1989).
12. Dorr, H. & Munnich, K. O. *Radiocarbon* **22**, 909-918 (1980).
13. Deines, P., Langmuir, D. & Harmon, R. S. *Geochim. cosmochim. Acta* **38**, 1147-1164 (1974).
14. Jenny, H. *Factors in Soil Formation* (McGraw-Hill, New York, 1941).
15. Birkeland, P. W. *Soils and Geomorphology* (Oxford University Press, New York, 1984).

ACKNOWLEDGEMENTS. We thank J. C. Barry, A. K. Behrensmeyer, W. Eckland, T. E. Fenton, P. Kneupper, D. Lewis, A. R. Mermut, J. Millar, C. G. Oviatt, W. G. Spaulding, J. L. Richardson, G. Smith, R. J. St. Arnaud, and W. Wehmueller for assistance in the field and in choosing the localities studied. This research was supported by the NSF and by the State of Nevada Nuclear Waste Project Office under a Department of Energy Grant from the U.S. Nuclear Waste Fund.

Crystal chemistry of phase B and an anhydrous analogue: implications for water storage in the upper mantle

L. W. Finger*, J. Ko†, R. M. Hazen*, T. Gasparik†, R. J. Hemley*, C. T. Prewitt* & D. J. Weidner†

* Geophysical Laboratory, Carnegie Institution of Washington, 2801 Upton Street NW, Washington, DC 20008, USA

† Department of Earth and Space Sciences, State University of New York, Stony Brook, New York 11794, USA

WATER locked in mineral phases within the Earth's mantle may be as significant for the development of life as water now found at the planet's surface. The Earth's original hydrosphere probably would not have survived a collision with a Mars-sized object, such as may have formed the Moon¹. The water now on the surface could have been replenished by further impacts of smaller planetesimals; however, at least a portion could have been stored in minerals deep in the mantle being released gradually through volcanic eruptions. This mechanism requires a stable high-pressure phase able to store water under mantle conditions. Of the several hydrous phases studied in the past, the material known as phase B has the highest density and is the only known hydrous form stable at pressures corresponding to depths of 400–500 km (refs 2, 3). Although phase B has been known for over twenty years, its crystal structure and crystal chemistry have remained an unsolved problem. Here we describe the crystal structures of phase B ($\text{Mg}_{12}\text{Si}_4\text{O}_{19}(\text{OH})_2$) and a chemically and structurally similar anhydrous form, AnhB ($\text{Mg}_{14}\text{Si}_5\text{O}_{24}$). These structures contain silicon in both fourfold and sixfold coordination; the silicon octahedra share all twelve edges with magnesium octahedra in a unique thirteen-cation cluster. Description of the structures of phases AnhB and B requires 18 and 40 atoms, respectively, demonstrating that high-pressure phases can have very complicated

crystal structures. The multitude of octahedral sites in these phases could lead to rather complicated fractionation behaviour during solid–solid or solid–liquid transitions in the presence of other octahedrally coordinated ions such as Al, Fe, Ti and Mn.

The stability of high-pressure hydrous phases has been known since the report² of preliminary data for what were named phases A, B and C. Phase B, which formed from gels with $\text{Mg}/\text{Si} = 3$ at pressures > 11 GPa, had the highest density of the new forms and seemed to be an interesting possibility for the mantle, whether hydrous or not. The determination^{3–5} of the unit cell, space group and infrared spectra of synthesized crystals of phase B confirmed the presence of hydroxyl groups; the structure, however, was not determined. The hydrous phase B was reported^{6,7} to be the second liquidus phase for natural lherzolite and synthetic compositions that were heated at a pressure of 20 GPa. No water was added to the capsules in these experiments and the hydrous phase seemed to persist at very high temperatures even though water usually enters the melt at relatively low temperature. Electron-microprobe results suggest⁸ that this high-temperature phase is actually anhydrous.

Single crystals of phase B were synthesized in a split-sphere anvil apparatus (USSA-2000) starting from a gel with $\text{Mg}/\text{Si} = 3$. The charge was heated to 1,200 °C at 12 GPa and cycled between 1,200 and 1,300 °C, a procedure that produced the largest grain size. Crystals selected from the run products have a monoclinic cell (see Table 1) similar to that reported in ref. 5. Some crystals demonstrate twinning, which we believe to be the result of a twofold operator parallel to [001], but, most did not. In addition, high-resolution transmission-electron-microscope photographs did not indicate the presence of stacking faults. The unit-cell geometry and space group are highly suggestive of a structure with some layers similar to that of MgO and a six-layer stacking sequence. In such a structure, the oxygen substructure would have cubic closest-packing arrangement, a situation confirmed by a Patterson synthesis. Because of this pseudosymmetry, a solution was not obtained by the standard direct methods.

Single crystals of the new phase, AnhB, were grown from a starting material consisting of synthetic forsterite that had been dried at 1,000 °C in argon immediately before the experiment; run conditions were 2,380 °C and 16.5 GPa for 2 h. The new phase formed large single crystals at the centre of the capsule, with quenched crystals in the hot end and $\beta\text{-Mg}_2\text{SiO}_4$ in the colder portion of the charge. Phase AnhB was formed by incongruent melting of $\beta\text{-Mg}_2\text{SiO}_4$ to AnhB plus melt, which segregated during the experiment. Electron-microprobe analysis gave an atomic Mg/Si ratio of 2.81 (9 analyses) and a total close to 100%. The lattice (see Table 1) also indicates a six-layer stacking sequence. Although this structure has pseudosymmetry, the solution was obtained by routine direct methods⁹ and difference electron-density calculations; the phase is isostructural¹⁰ with $\text{Mg}_{14}\text{Ge}_5\text{O}_{24}$. We have revised the choice of orthorhombic axes, however, to maintain the similarities between phases B and AnhB. To test for substitution of $\text{Mg} + \text{H}_2$ for Si, valence-bond sums¹¹ were calculated. All oxygen ions were fully bonded; furthermore, the Raman spectra (Fig. 1) showed none of the features associated with the presence of OH groups. An a - c projection of the structure is shown in Fig. 2. Not only is this the phase identified in ref. 8 but it is probably the high-temperature form identified as phase B in refs 6, 7. In addition, phase B has been found¹² to break down at high pressure at $\sim 1,000$ °C; above 22 GPa, the quenched sample consisted of one or more unknown phases. Phase AnhB, with some Fe substitution for Mg, would explain two of the three strongest X-ray lines found in that study. Another, as yet unidentified, phase is required to account for the third diffraction line. It is probable, however, that this phase is produced in a reaction that yields free water as a product.

Using the structure model for AnhB we obtained the solution of the hydrous form. The Raman spectrum (Fig. 1) indicates two distinct OH groups, which is consistent with the calculated

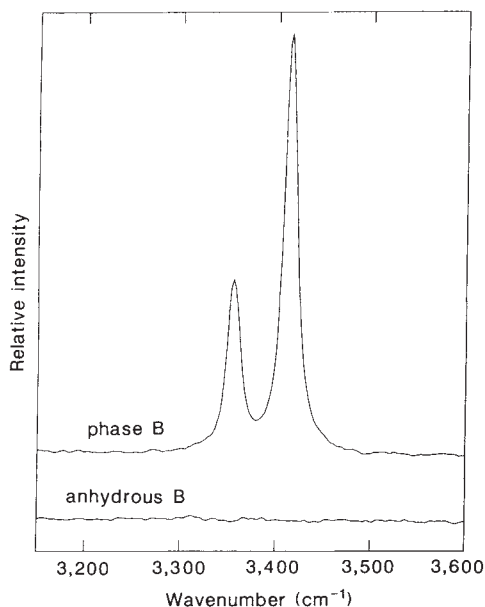


FIG. 1 Raman spectra of phase B (top curve) and anhydrous phase B (bottom curve) measured with 457.9-nm laser excitation (~ 25 mW) and 5-cm^{-1} resolution. The peaks in the phase B spectrum, which arise from OH-stretching vibrations, appear at $3,414$ and $3,356\text{ cm}^{-1}$, frequencies similar to those found⁴ in the infrared absorption spectrum.

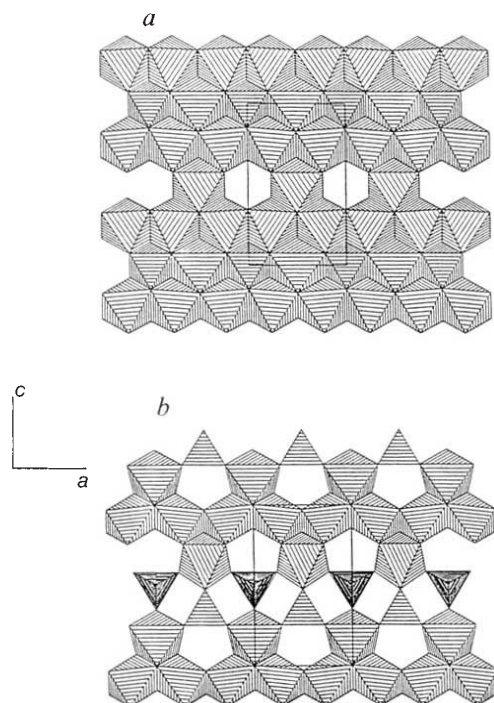


FIG. 2 Polyhedral²⁰ projection of the structure of ${}^{\text{VI}}\text{Mg}_{14}{}^{\text{IV}}\text{Si}_{14}\text{Si}_4\text{O}_{24}$ (AnhB) viewed parallel to (010). (Roman-numeral superscripts indicate the coordination). Structure consists of two types of layers repeated by symmetry. Magnesium is in six-fold coordination as in olivine or spinel; however, silicon is in mixed coordination. *a*, Defect (Mg,Si)O layer, with $-0.12 \leq y \leq 0.12$, similar to that of spinel²¹, but with fewer vacancies and with a smaller Si octahedron at the origin. There are two layers of this type per unit cell. *b*, Layer with $-0.30 \leq y \leq -0.07$ containing Mg octahedra and Si tetrahedra. There are four layers of this type per unit cell. The Si octahedron in *a* is capped by a cluster of Mg octahedra; thus all 12 edges of the Si octahedron are shared with Mg octahedra. The Si tetrahedra are placed directly over the hole in the first layer; thus, octahedra and tetrahedra share corners as in spinel rather than olivine; this layer is identical, however, to *b*-*c* layers of the olivine structure.

valence sums¹¹. These values are 1.2 and 1.3, respectively, compared to a value of 2.0 for a fully bonded ion. A nearby oxygen had a value of 1.6 indicating an asymmetric hydrogen bond. Trial positions were generated from these calculations and the parameters of all atoms were refined, using anisotropic thermal models for all non-hydrogen atoms. Figure 3 is an *a*-*c* projection of the structure. The hydrogen atoms lie 0.8–0.9 Å from the oxygen associated with a silicon tetrahedron.

The high density of these structures arises, in part, from total edge sharing between the silicon octahedron and the surrounding magnesium octahedra, which forms a thirteen-cation cluster. Unlike the lower-pressure forms olivine, clinohumite, humite, chondrodite and norbergite¹³, and the hydrous high-pressure form, phase A¹⁴, the new structures do not contain shared edges between tetrahedra and octahedra. In addition, only phases B and AnhB have silicon in octahedral coordination.

Other high-pressure forms of magnesium silicate may exist. The germanate form of AnhB has been described¹⁵ as a member of a homologous series containing variable ratios of olivine and defect rock-salt layers (in AnhB this ratio is 2:1). $\text{Mg}_{10}\text{Ge}_3\text{O}_{16}$, a stoichiometry that might occur in the silicate system has also been reported¹⁶. As shown in Fig. 3, phase B can be described as a 2:1 mixture of humite and defect rock salt. It is probable that the 1:1 form, which would have a hexagonal close-packed oxygen array with composition $\text{Mg}_{17}\text{Si}_5\text{O}_{26}(\text{OH})_2$, also exists. Other series based on clinohumite, chondrodite or norbergite layers mixed with the appropriate octahedral layers can be imagined. In addition as phase B can be related to AnhB by a

TABLE 1 Equivalent isotropic temperature factors and mean bond distances for cations of phases AnhB and B

	AnhB		B	
	B_{eq}	Mean M–O	B_{eq}	Mean M–O
Si 1	0.33 (2)*	1.804	0.26 (1)	1.816
Si 2	0.34 (1)	1.657	0.34 (2)	1.659
Si 3	0.35 (1)	1.625	0.35 (2)	1.655
Si 4			0.39 (2)	1.624
Mg 1	0.47 (2)	2.097	0.48 (2)	2.090
Mg 2	0.49 (2)	2.110	0.50 (2)	2.101
Mg 3	0.48 (1)	2.106	0.56 (3)	2.085
Mg 4	0.46 (1)	2.093	0.49 (2)	2.094
Mg 5	0.44 (1)	2.076	0.46 (2)	2.092
Mg 6	0.46 (1)	2.088	0.43 (2)	2.086
Mg 7			0.46 (2)	2.086
Mg 8			0.45 (2)	2.114
Mg 9			0.47 (2)	2.084
Mg 10			0.49 (2)	2.089
Mg 11			0.47 (2)	2.091
Mg 12			0.49 (2)	2.087
Mg 13			0.52 (3)	2.128

Diffraction experiments with Rigaku AFC-5 diffractometer, rotating-anode generator, Mo $\text{K}\alpha_1$ radiation, graphite monochromator, $\lambda = 0.7093$ Å, ω step scans to $2\theta = 60^\circ$, ambient pressure and temperature. Phase AnhB: $a = 5.868(1)$ Å; $b = 14.178(1)$ Å; $c = 10.048(1)$ Å; $V_{\text{cell}} = 835.9$ Å³. Space group $Pmcb$; $Z = 2$; relative molecular mass = 864.8; $\rho_{\text{calc}} = 3.435$ g cm⁻³; linear absorption coefficient $\mu_1 = 10.8$ cm⁻¹; range of transmission factors = 0.81–0.92; two octants measured; internal agreement 2.5%. Residuals $R = \sum ||F_o| - |F_c|| / \sum |F_o| = 0.040$, $R_w = \sum \sigma_F^2(F_o - F_c)^2 / \sum \sigma_F^2 F_o^2 = 0.025$ for all 1,337 independent data where F_o is the observed structure factor and F_c the calculated structure factor, $R = 0.029$, $R_w = 0.024$ for 1,132 data with $F_o > 2\sigma_F$; goodness of fit = 0.9 for anisotropic temperature factors. Phase B: $a = 10.588(2)$ Å; $b = 14.097(1)$ Å; $c = 10.073(1)$ Å; $\beta = 104.10(3)^\circ$; $V_{\text{cell}} = 1,458.4(3)$ Å³. Space group $P2_1/c$; $Z = 4$; relative molecular mass = 742.1; $\rho_{\text{calc}} = 3.368$ g cm⁻³; $\mu_1 = 10.3$ cm⁻¹; Range of transmission factors = 0.87–0.94; hemisphere of reciprocal space measured; internal agreement 3.5%. $R = \sum ||F_o| - |F_c|| / \sum |F_o| = 0.080$, $R_w = \sum \sigma_F^2(F_o - F_c)^2 / \sum \sigma_F^2 F_o^2 = 0.041$ for all 4,263 independent data; $R = 0.039$, $R_w = 0.056$ for 3,254 data with $F_o > 2\sigma_F$; goodness of fit = 1.0 for anisotropic temperature factors.

*Numbers in parentheses are the estimated standard deviations in the last decimal places.

'shearing' parallel to the *c* axis, one can propose other forms involving different slab widths or 'shear' offsets leading to further structures or complex intergrowths such as those found in the biopyrroboles¹⁷.

This study illustrates the utility of a large-volume high-pressure apparatus in growing single crystals of high-pressure phases that are suitable for crystal structure determination. This is important because the solution of these complicated structures would not be possible with powder X-ray diffraction methods. We note that very complex crystal structures can exist at high pressure in contradiction to the standard assumptions. For example, it has been proposed²² that the lower mantle is composed predominately of simple orthorhombic perovskites over the entire *P*-*T* range. Although there is no evidence to refute this claim, we emphasize that it has not yet been confirmed by *in situ* simultaneous high-pressure and high-temperature measurements.

This study also has implications regarding the storage and release of water in the upper mantle. Although these phases are more magnesian than any possible mantle model, mixture with stishovite would permit any desired Mg:Si value, with a high density. The assemblage B + stishovite with Mg/Si = 1.5 has 1.8% H₂O and an atmospheric-pressure density of 3.554, which is 9% denser than A + spinel + clinoenstatite of the same composition and 8% denser than clinohumite + stishovite¹⁸. Phase B is stable to 22 GPa at 1,000 °C (ref. 12), but transforms to AnhB at higher temperatures according to our interpretation of the data in ref. 6. This reaction, which might be B + spinel →

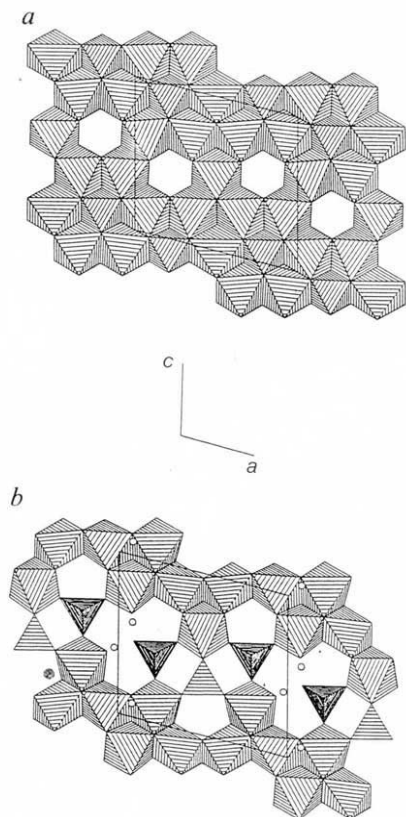


FIG. 3 Polyhedral projection of the structure of ${}^{\text{VI}}\text{Mg}_{12}{}^{\text{VI}}\text{Si}_{14}{}^{\text{IV}}\text{Si}_3\text{O}_{19}(\text{OH})_2$ (Phase B) viewed parallel to (010). (Roman-numeral superscripts indicate coordination.) This structure also has a two-layer form. *a*, The defect (Mg, Si)O layer, with $-0.12 \leq y \leq 0.12$, has a central vertical section identical to that of Fig. 2*a*; however, there is a 'sheared' relationship between adjacent strips in the hydrous form. Note that the silicon octahedron is no longer located at the origin. *b*, Layer with $-0.30 \leq y \leq -0.07$ containing Mg octahedra, Si tetrahedra, and hydrogens indicated by small circles. This layer is identical to *b*-*c* layers of the humite structure.

$\text{AnhB} + \text{H}_2\text{O}$ or $\text{B} + \text{stishovite} \rightarrow \text{AnhB} + \text{H}_2\text{O}$, could be a mechanism for storage and release of large volumes of water. A possible scenario is that during the early history of the mantle, a magma ocean might first crystallize garnet, then AnhB. On cooling in the presence of water, the reaction would proceed to the left, storing the water in the solids. The difficulty with this scenario is that the hydrous phase B has not been shown to be stable in realistic mantle compositions; its phase relations are little known. We believe that there has been insufficient water present in most experiments to stabilize hydrous phase B. Further work is certainly required to clarify details of the possible phase relations.

Even if high-pressure hydrous silicates were not involved in the formation of primordial surface water, there is circumstantial evidence for their importance in maintaining a hydrosphere. Those planets too small to stabilize phase B in their mantles (Moon, Mercury) are completely dry¹. The roughly 20-GPa pressures at the core-mantle boundary of Mars would limit the volume of phase B and its surface water has not been replaced. Venus is more complicated. Phase B should be abundant but at a greater depth than on Earth. The quantity of water in the atmosphere will be affected by the efficiency of volcanism in tapping this source and by the extent to which atmospheric water is photodissociated and H_2 lost¹⁹. □

3. Akaogi, M. & Akimoto, S. *J. geophys. Res.* **84**, 6944-6948 (1980).
4. Akaogi, M. & Akimoto, S. *Phys. Chem. Miner.* **13**, 161-164 (1986).
5. Akimoto, S. & Akaogi, M. in *Materials Science of the Earth's Interior* (ed. Sunagawa, I.) 477-480 (Terra Scientific, Tokyo, 1984).
6. Kato, T. & Kumazawa, M. *Geophys. Res. Lett.* **13**, 181-184 (1986).
7. Kato, T. & Kumazawa, M. *Geophys. Res. Lett.* **12**, 534-535 (1985).
8. Herzberg, C. T. & Gasparik, T. *Eos* **70**, 484 (1989).
9. Gilmore, G. J. *J. appl. Crystallogr.* **17**, 42-46 (1984).
10. von Dreele, R. B., Bless, P. W., Kostiner, E. & Hughes, R. E. *J. Solid St. Chem.* **2**, 612-618 (1970).
11. Brown, I. D. in *Structure and Bonding in Crystals II* (eds O'Keeffe, M. & Navrotsky, A.) 1-30 (Academic, New York, 1981).
12. Liu, L. *Phys. Earth planet. Inter.* **42**, 255-262 (1986).
13. Papike, J. J. & Cameron, M. *Rev. Geophys. Space Phys.* **14**, 37-80 (1976).
14. Horiuchi, H., Morimoto, N., Yamamoto, K. & Akimoto, S. *Am. Miner.* **64**, 593-598 (1979).
15. Barbier, J. *Acta crystallogr.* **B43**, 422-429 (1987).
16. Barbier, J. *J. Solid St. Chem.* **68**, 52-60 (1987).
17. Veblen, D. R. & Buseck, P. R. *Am. Miner.* **64**, 687-700 (1979).
18. McGetchen, T. R., Silver, L. T. & Chodos, A. A. *J. geophys. Res.* **75**, 255-259 (1970).
19. Matsui, T. & Abe, Y. *Nature* **322**, 526-528 (1986).
20. Fisher, R. X. *J. appl. Crystallogr.* **18**, 258-262 (1984).
21. Morimoto, N., Tokonami, M., Watanabe, M. & Koto, K. *Am. Miner.* **59**, 475-485 (1974).
22. Lin, L. *Geophys. Res. Lett.* **9**, 124-126 (1982).

ACKNOWLEDGEMENTS. We thank D. Veblen for transmission-electron-microscope examinations of phase B. The synthesis of single crystals was performed in the Stony Brook High-Pressure Laboratory, which is jointly supported by NSF and the State University of New York, and the X-ray diffraction data were obtained at the Geophysical Laboratory, which is supported by NSF and by the Carnegie Institution of Washington.

Ecosystem-level patterns of primary productivity and herbivory in terrestrial habitats

S. J. McNaughton, M. Oesterheld, D. A. Frank & K. J. Williams

Biological Research Laboratories, Syracuse University, Syracuse, New York 13244-1220, USA

ECOSYSTEMS are structurally organized as food webs within which energy is transmitted between trophic levels and dissipated into the environment. Energy flow between two trophic levels is given by the amount of production at the lower level and by the proportion of production that is consumed, assimilated and respired at the higher level. Considerable evidence indicates that food-web structure varies predictably in different habitats¹⁻⁵, but much less is known about quantitative relationships among food web fluxes. Many of the energetic properties of herbivores in African game parks are associated with rainfall and, by inference, with net primary productivity^{6,7}. Respiratory costs per unit production at the consumer trophic level are higher for homeotherms than for heterotherms⁸. Plant secondary chemicals affect herbivore dietary choices^{9,10} and the allocation of plant resources to those chemicals varies with resource availability¹¹. How these phenomena are translated into ecosystem fluxes is unknown. We present evidence that herbivore biomass, consumption and productivity are closely correlated with plant productivity, suggesting that the latter is a principal integrator and indicator of functional processes in food webs.

A data base was compiled, using sets of data reported in the literature, which combined measurements of herbivore biomass, consumption and/or net secondary productivity with measures of net primary productivity^{6,7,12-39}. Unpublished estimates of consumption and net primary productivity in unmanaged African grasslands, from a study described earlier⁴⁰, were also included. Cases in which herbivore trophic-level properties were estimated from primary productivity were excluded. Values of herbivore biomass, consumption and net secondary productivity rarely include all populations of herbivores in an ecosystem. Consequently, data were included only if the preponderance of the herbivore level was accounted for in the original study. Only processes occurring above ground were included, as comprehensive below-ground measurements were too few in numbers for quantitative analysis. Fluxes expressed in the literature as units

Received 5 June; accepted 8 August 1989.

1. Newson, H. E. & Taylor, S. R. *Nature* **338**, 29-34 (1989).

2. Ringwood, A. E. & Major, A. *Earth planet. Sci. Lett.* **22**, 130-133 (1967).

of mass were converted to energy using standard conversion factors⁴¹⁻⁴³. Values of net primary production, net secondary production and consumption are presented in $\text{kJ m}^{-2} \text{yr}^{-1}$, and values of herbivore biomass are in kJ m^{-2} . Net primary productivity ranged from ~ 125 to $\sim 29,000 \text{ kJ m}^{-2} \text{yr}^{-1}$. Wide ranges of ecosystems and primary productivities were thus encompassed by the assembled data.

There was a positive association between primary productivity and all three energetic properties of herbivore communities, with different types of ecosystems segregating along common lines (Fig. 1). Herbivore biomass (B) was related (coefficient of determination $r^2 = 0.579$, $P < 0.00001$, degrees of freedom (d.f.) = 49) to net primary productivity (NAP) by

$$\log B = 1.52 (\log \text{NAP}) - 4.79$$

where log indicates common logarithms. Consumption (C) and NAP were related ($r^2 = 0.367$, $P < 0.00001$, d.f. = 67) by

$$\log C = 1.38 (\log \text{NAP}) - 2.32$$

Finally, net secondary productivity (NSP) was related ($r^2 = 0.364$, $P < 0.0001$, d.f. = 34) to NAP by

$$\log \text{NSP} = 1.10 (\log \text{NAP}) - 3.27$$

These data indicate that net energy flux into the food-web base

through primary producers is indicative of the three major energetic properties of the trophic level directly above.

Inspection of the goodness of fit between consumption and net primary productivity, revealed that forests fell below the line and grasslands above. Dominance by heterotherm herbivores in forests and homeotherms in grasslands, and the greater prevalence of secondary chemicals in trees than in grasses¹¹, were initially accepted as explanation of this pattern. Much of the net productivity of forests, however, is confined to wood, a poor quality food⁴⁴ largely unavailable to most herbivores. Therefore, we re-examined the relationship, confining net productivity to foliage (NFP). Then, for the larger number of studies in which foliage production and consumption were concurrently measured in forests, the relationship was (Fig. 2)

$$\log C = 2.04 (\log \text{NFP}) - 4.80$$

($r^2 = 0.594$, $P < 0.00001$, d.f. = 73). Restricting primary productivity data to foliage, therefore, decreased the unexplained variance substantially by bringing forested ecosystems into line with the patterns across other ecosystems.

We infer that secondary chemicals and differences in the basic physiologies of consumers are much less important influences upon consumption at the ecosystem level than they are at the individual plant and population levels. Other than contributing to an essentially unavailable food class, namely wood⁴⁴, secondary chemicals appear to be more important as regulators of what herbivores eat and the types of herbivores prevalent in different ecosystems than as regulators of consumption at the ecosystem level.

All major energetic parameters of the herbivore community—biomass, incoming energy (consumption), and net energy change (net secondary productivity)—were closely associated with net energy flux into the food web. Those relationships based on best fit equations, however, were different. Herbivore biomass increased as a power of net primary productivity as the slope was significantly greater than 1 ($t = 2.80$, $P < 0.01$). The slope of consumption was also greater than 1 ($t = 1.71$, $P < 0.05$), particularly when primary productivity was restricted to foliage ($t = 5.26$, $P < 0.001$). Net secondary production, however, increased linearly with net primary production, with a slope indistinguishable from 1 ($t = 0.38$, P , not significant). This indicates that highly productive ecosystems sustain a larger level of herbivory per unit of primary production than unproductive systems. This larger level of herbivory, however, is accompanied by a lower efficiency in secondary production per unit of consumption, and, as a result, the relationship between net secondary and net primary production is linear.

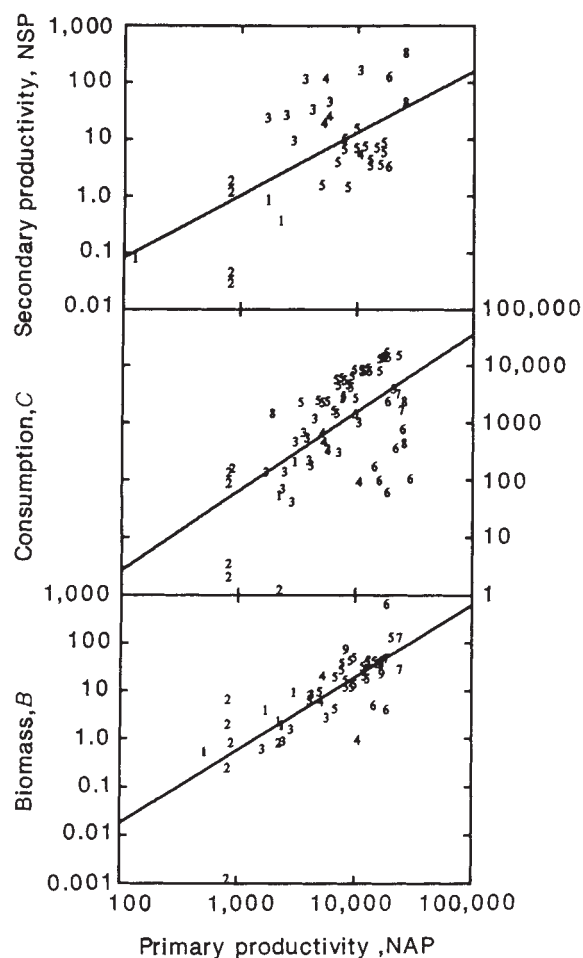


FIG. 1 Relationship between net above-ground primary productivity (NAP, abscissa) and herbivore biomass, consumption, and net secondary productivity (NSP). Units are $\text{kJ m}^{-2} \text{yr}^{-1}$, except for biomass which is kJ m^{-2} . Key to type of ecosystem: 1, desert; 2, tundra; 3, temperate grassland; 4, temperate successional old field; 5, unmanaged tropical grassland; 6, temperate forest; 7, tropical forest; 8, salt marsh; and 9, agricultural tropical grassland. The data set consisted of 51 points for B , 69 for C and 35 for NSP.

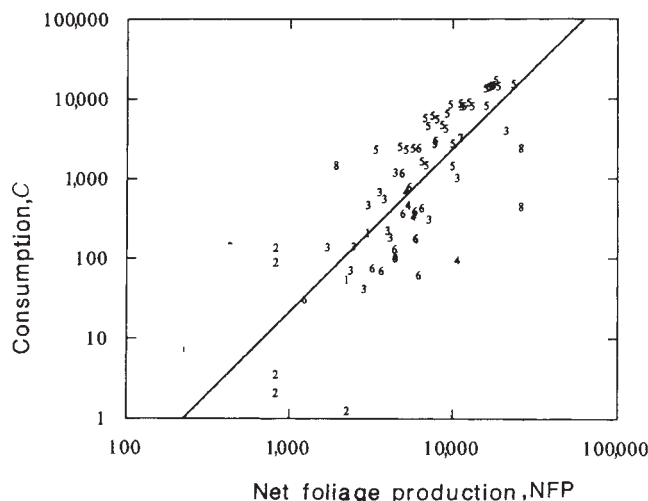


FIG. 2 Relationship between net foliage production (NFP) and consumption (C) by herbivores. Units and ecosystem symbols, as in Fig. 1.

The relationship between consumption and net foliage productivity was heavily influenced, in its upper limit, by a prevalence of data from African game parks, known to support among the highest levels of herbivory yet quantified⁴⁵. It would be desirable to have a greater representation of other systems believed to be highly productive, such as tropical rain forests. Nevertheless, it is recognized that the levels of herbivory are variable but can be extremely high in tropical forest ecosystems^{46,47}. High levels of folivory and rapid recycling of nutrients through herbivore wastes, therefore, may be a general feature of highly productive ecosystems⁴⁸.

We believe that it would be incorrect to infer from our results that biomass, consumption and production at higher trophic levels are merely a passive consequence of the incorporation of energy and chemicals into the producer level⁴⁰. Higher trophic levels can often have dramatic effects upon food-web structure⁴⁹ and on both growth form⁵⁰ and productivity of plants⁴⁰. It is useful, we believe, to regard net primary productivity as an integrative ecosystem variable that both influences, and is a reflection of, processes occurring at other trophic levels, some widely removed from primary producers⁵¹.

The relationships documented here between energetic input at the base of the food web (net primary production) and major energetic components of the herbivore trophic level directly above have three major implications for ecosystem science. First, measuring net primary production, which is generally more feasible than measuring processes at the consumer trophic level, and may even be practicable using satellite imagery⁵²⁻⁵⁴, can provide extensive information on whole-system fluxes. The global approach of the International Geosphere-Biosphere program could provide extensive information on general ecosystem processes if the relationships documented here are widely applicable. Second, as food-web structure is related in predictable ways to habitat¹⁻⁵, it may be possible to relate structural organizations of food webs to their functional properties in a more integrated, straightforward and predictive fashion. Finally, although our results were confined to only two trophic levels, their regularity indicates that similar patterns will apply to other trophic levels. As, for example, the foliage unconsumed by herbivores will flow into decomposer food webs, the relationship documented here between consumption by herbivores and net primary production indicates that the relative importance of the direct flow to detritus decreases as ecosystem productivity increases. Using the best fit equation, the predicted proportion of annual foliage production processed directly by decomposers, without the intervention of consumers, would decrease from >99% at a net foliage productivity of 500 kJ m⁻² yr⁻¹ to <50% at a net foliage productivity of 20,000 kJ m⁻² yr⁻¹. □

22. Haukioja, E. & Koponen, S. in *Fennoscandian Tundra Ecosystems. Part 2. Animals and Systems Analysis* (ed. Wielgolaski, F. E.) 182-188 (Springer, Heidelberg, 1975).
23. Sinclair, A. R. E. *J. Anim. Ecol.* **44**, 497-520 (1975).
24. Wielgolaski, F. E. in *Fennoscandian Tundra Ecosystems* (ed. Wielgolaski, F. E.) 300-326 (Springer, Heidelberg, 1975).
25. Grier, C. G. & Logan, R. S. *Ecol. Monogr.* **47**, 373-400 (1977).
26. Nielson, O. *Oikos* **31**, 273-279 (1978).
27. Coupland, R. T. & Van Dyne, G. M. in *Grassland Ecosystems of the World; International Biological Programme* (ed. Coupland, R. T.) 97-106 (Cambridge University Press, 1979).
28. French, N. R., Steinhilber, R. K. & Swift, D. M. in *Perspectives in Grassland Ecology* (ed. French, N.) 59-88 (Springer, New York, 1979).
29. Scott, J. A., French, N. R. & Leatham, J. W. in *Perspectives in Grassland Ecology* (ed. French, N.) 89-106 (Springer, New York, 1979).
30. Sims, P. L. & Coupland, R. T. in *Grassland Ecosystems of the World* (ed. Coupland, R. T.) 49-72 (Cambridge University Press, 1979).
31. Batzli, G. O., White, R. G., MacLean, S. F., Pitelka, F. A. & Collier, B. D. in *An Arctic Ecosystem. The Coastal Tundra at Barrow, Alaska* (eds Brown, J., Miller, P. C., Tieszen, L. L. & Bunnell, F. L.) 335-410 (Dowden Hutchinson Ross, Stroudsburg, 1980).
32. DeAngelis, D. L., Gardner, R. H. & Shugart, H. H. in *Dynamic Properties of Forest Ecosystems* (ed. Reichle, D. E.) 567-672 (Cambridge University Press, New York, 1981).
33. Pfeiffer, W. J. & Wiegert, R. G. in *The Ecology of a Salt Marsh* (eds Pomeroy, L. R. & Wiegert, R. G.) 87-112 (Springer, New York, 1981).
34. Schowalter, T. D., Webb, J. W. & Crossley, D. A. *Ecology* **62**, 1010-1019 (1981).
35. Leigh, E. G. & Windsor, D. M. in *The Ecology of a Tropical Forest. Seasonal Rhythms and Long-term Changes* (eds Leigh, E. G., Rand, A. S. & Windsor, D. M.) 111-122 (Smithsonian Institution, Washington, 1982).
36. Ohmart, C. P., Stewart, L. G. & Thomas, J. R. *Oecologia* **59**, 322-330 (1983).
37. Cargill, S. M. & Jefferies, R. L. *J. appl. Ecol.* **21**, 669-686 (1984).
38. Coughenour, M. B. *et al. Science* **230**, 619-625 (1985).
39. Bunderson, W. T. in *Rangelands: a Resource Under Siege* (eds Joss, P. J., Lynch, P. W. & Williams, O. B.) 397-401 (Australian Academy of Sciences, Canberra, 1986).
40. McNaughton, S. J. *Ecol. Monogr.* **55**, 259-294 (1985).
41. Golley, F. B. *Ecology* **42**, 581-584 (1961).
42. Schroeder, L. A. *Oecologia* **28**, 261-267 (1977).
43. Whitfield, D. W. A. in *Truelove Lowland, Devon Island, Canada: A high Arctic Ecosystem* (ed. Bliss, L. C.) 607-620 (University of Alberta Press, Edmonton, 1977).
44. Loehle, C. *Can. J. Forest Res.* **18**, 209-233 (1988); *Vegetatio* **77**, 65-78 (1988).
45. McNaughton, S. J. & Georgiadis, N. J. A. *Rev. Ecol. Syst.* **17**, 39-65 (1986).
46. Coley, P. D. *Ecol. Monogr.* **53**, 209-233 (1983); *Ecology* **64**, 426-433 (1983).
47. Brown, B. J. & Ewel, J. J. *Ecology* **68**, 108-116 (1987).
48. Ruess, R. W. & McNaughton, S. J. *Oecologia* **63**, 331-337 (1984).
49. Paine, R. T. *Am. Nat.* **100**, 65-75 (1966); *J. Anim. Ecol.* **49**, 667-685 (1980).
50. McNaughton, S. J. & Sabuni, G. A. in *Plant Form and Vegetation Structure* (eds Werger, M. J. A., van der Aart, P. J. M., Durning, H. J. & Verhoeven, J. T. A.) 339-354 (SPB, The Hague, 1988).
51. McNaughton, S. J., Ruess, R. W. & Seagle, S. W. *BioScience* **38**, 794-800 (1988).
52. Tucker, C. J., Townshend, J. R. G. & Goff, T. E. *Science* **227**, 369-375 (1985).
53. Space Science Board *Remote Sensing of the Biosphere* (National Academy Press, Washington, 1986).
54. Earth System Sciences Committee, NASA Advisory Council *Earth System Science. A Closer View*. (NASA, Washington, 1988).

ACKNOWLEDGEMENTS. This research was stimulated by a request to participate in the Fourth Cary Conference at the Institute of Ecosystem Studies, Mary Flagler Cary Arboretum of The New York Botanical Garden. Financial support was from the NSF.

The secreted form of the Alzheimer's amyloid precursor protein with the Kunitz domain is protease nexin-II

Tilman Oltersdorf, Lawrence C. Fritz, Dale B. Schenk, Ivan Lieberburg, Kelly L. Johnson-Wood, Eric C. Beattie, Pamela J. Ward, Russell W. Blacher, Harry F. Dovey & Sukanto Sinha*

Athena Neurosciences, Inc., 800 F Gateway Boulevard, South San Francisco, California 94080, USA

THE A4 protein (or β -protein) is a 42- or 43-amino-acid peptide¹ present in the extracellular neuritic plaques in Alzheimer's disease and is derived from a membrane-bound amyloid protein precursor (APP). Three forms of APP have been described²⁻⁷ and are referred to as APP695, APP751 and APP770, reflecting the number of amino acids encoded for by their respective complementary DNAs. The two larger APPs contain a 57-amino-acid insert with striking homology to the Kunitz family of protease inhibitors⁵⁻⁷. Here we report that the deduced amino-terminal sequence of APP is identical to the sequence of a cell-secreted protease inhibitor, protease nexin-II (PN-II)⁸. To confirm this finding, APP751 and APP695

* To whom correspondence should be addressed.

Received 22 May; accepted 4 August 1989.

1. Cohen, J. *Nature* **270**, 165-166 (1977).
2. Pimm, S. L. *Food Webs* (Chapman and Hall, London, 1982).
3. Briand, F. *Ecology* **64**, 253-263 (1983).
4. Briand, F. & Cohen, J. *Nature* **307**, 264-267 (1984). *Science* **238**, 956-960 (1987); **243**, 239-240 (1989).
5. Moore, J. C., Walter, D. E. & Hunt, H. W. *Science* **243**, 238-239 (1989).
6. Coe, M. J., Cumming, D. H. & Phillipson, J. *Oecologia* **22**, 341-354 (1976).
7. East, R. *Afr. J. Ecol.* **22**, 245-270 (1984).
8. McNeill, S. & Lawton, J. H. *Nature* **225**, 472-474 (1970).
9. Whittaker, R. H. & Feeny, P. P. *Science* **171**, 757-770 (1971).
10. Rosenthal, G. A. & Janzen, D. H. (eds), *Herbivores* (Academic, New York, 1979).
11. Coley, P. D., Bryant, J. P. & Chapin, F. S. III *Science* **230**, 895-899 (1985).
12. Smalley, A. E. *Ecology* **41**, 672-677 (1960).
13. Golley, F. B., Odum H. T. & Wilson, R. F. *Ecology* **43**, 9-19 (1962).
14. Bray, R. J. *Ecology* **45**, 165-167 (1964).
15. Menhinick, E. F. *Ecol. Monogr.* **37**, 255-273 (1967).
16. Wiegert, R. G. & Evans, F. C. in *Secondary Productivity of Terrestrial Ecosystems* (ed. Petrusiewicz, K.) 499-518 (Polish Academy of Sciences, Krakow, 1967).
17. Chew, R. M. & Chew, A. E. *Ecol. Monogr.* **40**, 1-21 (1970).
18. Odum, H. T. *A Tropical Rain Forest: A Study of Irradiation and Ecology at El Verde, Puerto Rico* (Division of Technical Information and Atomic Energy Commission, Springfield, 1970).
19. Van Hook, R. I. *Ecol. Monogr.* **41**, 1-26 (1971).
20. Gosz, J. R., Holmes, R. T., Likens, G. E. & Bormann, F. H. *Scient. Am.* **238**, 93-102 (1978).
21. Harris, W. F., Sollins, P., Edwards, N. T., Dinger, B. E. & Shugart, H. H. in *Productivity of World Ecosystems* (eds Reichle, D. E., Franklin, J. F. & Goodall, D. W.) 116-122 (National Academy of Sciences, Washington, 1975).

PN-II	1	Leu	Glu	Val	Pro	Thr	Asp	Gly	Asn	Ala	Gly	10
APP	18	Leu	Glu	Val	Pro	Thr	Asp	Gly	Asn	Ala	Gly	29
PN-II	12	Leu	Ala	Glu	Pro	Gln	Ile	Ala	Met	Phe	Cys	21
APP	31	Leu	Ala	Glu	Pro	Gln	Ile	Ala	Met	Phe	Cys	40
PN-II	23	Arg	Leu	Asn	Met	Phe	Met	Asn	Val	Gln	Asn	32
APP	42	Arg	Leu	Asn	Met	His	Met	Asn	Val	Gln	Asn	52

FIG. 1 Alignment of the N-terminal sequence of PN-II with the deduced N-terminal sequence (residues 18–50 of primary transcript) of APP^{9,10}. The disagreement at residue 27 of PN-II is probably due to a mis-determination during Edman degradation.

cDNAs were over-expressed in the human 293 cell line, and the secreted N-terminal extracellular domains of these APPs were purified to near homogeneity from the tissue-culture medium. The relative molecular mass and high-affinity binding to dextran sulphate of secreted APP751 were consistent with that of PN-II. Functionally, secreted APP751 formed stable, non-covalent, inhibitory complexes with trypsin. Secreted APP695 did not form complexes with trypsin. We conclude that the secreted form of APP with the Kunitz protease inhibitor domain is PN-II.

The sequence of APP was compared by computer-assisted methods with sequences in the Protein Identification Resource (PIR) protein-sequence database. Surprisingly, amino-acid residues 18 to 50 of APP are identical to the N-terminal 33-amino-acid sequence of PN-II (ref. 8; Fig. 1) except at position 27 of PN-II; the residue at this position is a phenylalanine in

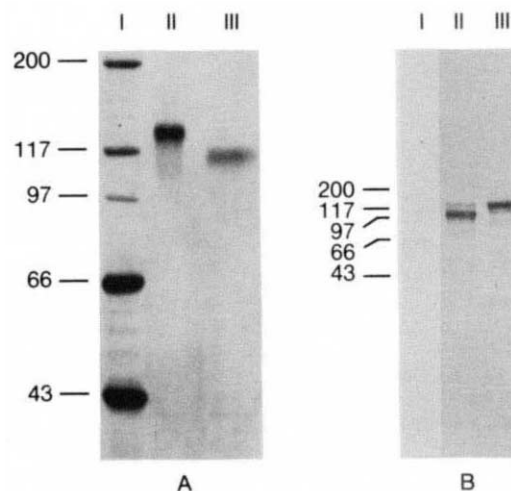
contrast to the histidine at the corresponding position of APP. This difference is not definite, however, because the identification of residue 27 was in doubt in the original publication of the sequence of PN-II. The N-terminal 17 amino acids of the APP sequence, which precede the PN-II homology region, correspond exactly to the presumed signal sequence, which is cleaved off during translocation into the endoplasmic reticulum^{9,10}.

Protease nexins are a heterogeneous group of serine protease inhibitors secreted by a number of cell types in culture (see refs 11–13 for reviews). PN-II was first isolated by its property of forming complexes with epidermal growth factor-binding protein, EGF-BP, a pro-EGF processing serine esterase^{14–17}. PN-II was subsequently purified to homogeneity from conditioned medium of primary human fibroblasts⁸, and shown to additionally form complexes with trypsin and with the gamma subunit of nerve growth factor (NGF).

To further establish the relationship between PN-II and APP we co-transfected 293 transformed human-embryonic-kidney cells with pSV2-neo and cDNAs coding for APP695 and APP751 under the control of the human cytomegalovirus promoter/enhancer, which were then over-expressed as in stable clones¹⁸. The secreted form of APP was detected in tissue-culture medium by western blotting, using an affinity-purified polyclonal antibody raised against a bacterial fusion protein containing the extracellular part of APP from amino-acid residues 444 to 592 (anti-secreted APP) (Fig. 2). The apparent relative molecular masses (M_r) of secreted APP695 and APP751, determined by SDS-PAGE under reducing conditions, were ~110,000 (110 K) and 124,000 (124 K), respectively (Fig. 2a; Bio-Rad standards, high range; Amersham protein standards led to estimates of M_r of 9–10 K less). The reported M_r of PN-II under these conditions is 112 K (ref. 8). As PN-II has been shown to bind dextran sulphate-Sepharose with high affinity, we used dextran sulphate-Sepharose chromatography, followed by DEAE-Sepharose

FIG. 2 a, SDS-PAGE analysis of purified secreted APP695 and APP751. Lane I, Bio-Rad high-range protein standards, 250 ng of each (myosin, 200 K; beta-galactosidase, 117 K; phosphorylase b, 97 K; BSA, 66 K; and ovalbumin, 43 K; lane II, secreted APP751 (175 ng); lane III, secreted APP695 (150 ng). b, Western blot of purified secreted APP695 and APP751. Lane I, Bio-Rad high range protein standards; lane II, secreted APP695; lane III, secreted APP751.

METHODS. Purification of secreted APPs. 293 cells stably transfected with either APP695 or APP751 cDNAs were grown in serum-free F12/Dulbecco's minimum essential medium (1:1) supplemented with 50 μ M ethanolamine, 10 nM sodium selenite and 1 mM sodium pyruvate. Conditioned medium (500 ml) was collected from each cell type, then centrifuged and filtered. The clear medium was then loaded on to a 1.6 $\times$ 15-cm dextran-Sepharose column (prepared as described previously⁹), then washed with 50-column volumes of 20 mM potassium phosphate buffer, pH 7.4 and 75 mM NaCl. The column was then eluted with a linear gradient to 1.2 M NaCl in the same buffer. Fractions containing secreted APPs were identified by western blotting using affinity purified rabbit antisera raised to a bacterially expressed fusion protein containing the portion of APP amino-acid sequence from residue 444–592 (restriction sites *Bam*HI to *Bgl*II). This cDNA fragment was inserted in the *Bam*HI restriction site of plasmid pEx $\times$ 10-mer (ref. 28), and a stop codon was introduced by making the plasmid linear using the *Xba*I site, filling in with Klenow polymerase, and religating the resulting blunt ends. The resulting plasmid, pBX5, was used to obtain a fusion protein with a 100 amino-acid bacteriophage MS2 polymerase N-terminal leader sequence by induction of the bacteriophage λ P₁ promoter contained in these plasmids, and partially purified by urea extraction as described previously²⁸. Rabbit antisera raised against this fusion protein (anti-pBX5) were affinity purified by first removing non-relevant IgG over a column containing immobilized bacterially expressed pEx10-mer (containing the MS-2 leader peptide), then by binding to, and eluting from, a second column containing immobilized partially purified pBX5. Pooled fractions from the dextran-Sepharose eluate were then adjusted to the conductivity of 10 mM Tris buffer, pH 8, and NaCl (50 mM), by dilution with 10 mM Tris, pH 8, and then loaded onto a 1.6 $\times$ 10-cm DEAE-Sepharose column equilibrated with this buffer. The column was eluted with a linear gradient to 1.2 M NaCl, and fractions containing



secreted APPs were again identified by western blotting and then pooled. Protein concentrations in the pooled fractions were estimated by the BCA protein assay (Pierce). Protein (150–175 ng) from each pool was analysed by SDS-PAGE, on a 12 $\times$ 18-cm acrylamide gel (10%), after denaturing in Laemmli sample buffer with DTT (100 mM). After electrophoresis, the gel was silver-stained²⁹. For the western blot analysis, ~150 ng of purified secreted APPs were electrophoresed on a 12% acrylamide 6 $\times$ 9-cm SDS-PAGE, in a Bio-Rad Mini-Protean gel system, and the protein transferred to Immobilon (Millipore) membrane. The membrane was then incubated with anti-pBX5 (1:500), and the western blot developed using goat anti-rabbit IgG-alkaline phosphatase (Bio-Rad). The sensitivity of the western blot technique was shown by the detection of contaminating secreted APP751 in the purified secreted APP695 fraction, which is present at a level below the detection limit of the silver-staining method in a, lane III.

chromatography, to purify secreted APPs from serum-free conditioned medium of transfected cells (Fig. 2). The analysis of the two purified secreted APPs by SDS-PAGE silver staining and western blotting is shown in Fig. 2.

Aliquots of the purified secreted APPs were then tested for their ability to inhibit the esterolytic activity of pancreatic trypsin (Fig. 3a). Secreted APP751 potentially inhibited trypsin, whereas secreted APP695 had no discernible inhibitory activity.

HPLC-gel exclusion chromatography (GEC) was used to further confirm that a quantitative 1:1 complex formation of trypsin with secreted APP751 occurs in physiological buffer conditions. Porcine pancreatic trypsin was radio-iodinated by the lactoperoxidase-glucose oxidase method¹⁹, and affinity purified on a column of Aprotinin-Sepharose²⁰. Purified secreted APP751 was incubated with a stoichiometric mixture

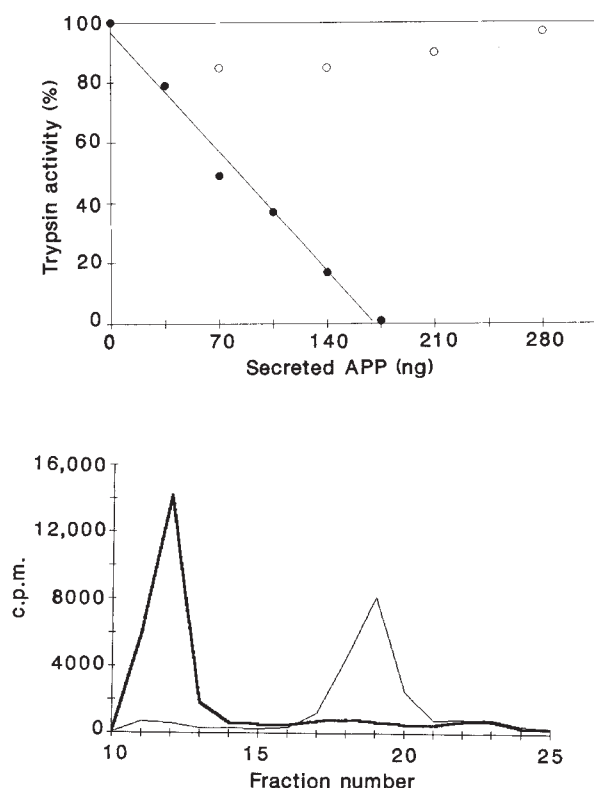


FIG. 3 a, Inhibition of the amidolytic activity of bovine pancreatic trypsin by secreted APPs. Aliquots (40 ng) of bovine pancreatic trypsin were incubated with increasing amounts of secreted APPs (APP751, 0–175 ng; APP695, 0–280 ng) in a total volume of 100 μ l, containing 0.1 M HEPES, pH 7.25, and 1 mM CaCl_2 . After 5 min, residual trypsin esterolytic activity was assayed by diluting 10 μ l of the mixture into 140 μ l of assay solution containing 0.25 mM Z-valyl-glycyl-arginyl-p-nitroanilide, in 0.1 M HEPES, pH 7.25, and 1 mM CaCl_2 , and the rate of hydrolysis of the substrate monitored by the increase in absorbance at 405 nm. Kinetic assays were done in microtitre plate wells in a Molecular Devices V^{max} machine. Rate values were converted to 'percent trypsin activity' relative to the rate of hydrolysis of substrate by trypsin in the absence of addition of secreted APP. The inhibition of trypsin activity by secreted APP751 (●) could be fitted to a linear regression line, and showed virtually 1:1 (molar ratio) inhibition of trypsin activity by secreted APP751. By contrast, secreted APP695 (○) had no significant effect on trypsin activity. b, HPLC-GEC of complexes of trypsin with secreted APP751. Trypsin (40 ng) was mixed with 1 ng radio-iodinated trypsin (50,000 c.p.m.) then mixed either with 20 mM HEPES, pH 7.25, and 0.5 M NaCl to a final volume of 100 μ l, or with 150 ng purified secreted APP751 in the same buffer. The mixtures were then injected onto a Bio-Rad TSK125 column, 0.78 $\times$ 25 cm, equilibrated with the same buffer at a flow rate of 0.5 ml min⁻¹. Fractions (0.5 ml) were collected and counted for [¹²⁵I] activity directly in a Beckman Gamma Counter. Free trypsin, fine line; APP751, bold line.

of cold trypsin and 125 I-labelled trypsin, and the mixture was then resolved on an HPLC-GEC column equilibrated with phosphate HEPES (20 mM pH 7.4) and NaCl (150 mM 0.5 M) (Fig. 3b). In the absence of added APP751, free iodinated trypsin eluted with a peak in fraction 19. Pre-incubation with secreted APP751 led to virtually quantitative peak elution in fraction 12, corresponding to an apparent higher M_r and indicating stable complex formation.

To test whether the inhibition of trypsin by secreted APP751 was accompanied by the formation of complexes stable to denaturing SDS-PAGE, as has been claimed for complexes of PN-II with trypsin and the other arginine esterases⁸, purified secreted APP751 was briefly incubated with a sub-stoichiometric amount of trypsin, mixed in a 1:1 ratio with twofold concentrated Laemmli sample buffer²¹, heated at 60 °C for 5 min, then analysed by non-reducing SDS-PAGE. We found that complexes treated in this way were not stable to denaturing SDS-PAGE, because the trypsin and the secreted APP751 migrated with their usual independent mobilities (Fig. 4, lane 3B). We were also unable to show SDS stability using a Bistris-Bicine buffer SDS-PAGE system²², similar to a Monotris-Bistris buffer system used by Knauer *et al.*¹⁴, which was claimed to enhance SDS stability of PN-II-esterase complexes. When the trypsin-secreted APP751 mixtures were incubated in Laemmli buffer at room temperature, however, a higher M_r complex of trypsin with secreted APP751 (Fig. 4, lane 3A) was detected. With purified secreted APP695 we were unable to detect any complex formation; in fact, the brief incubation with trypsin before electrophoresis led to partial fragmentation of this form of the secreted APP (Fig. 4, lanes 4A and B). Only the 751 form, therefore, is capable of inhibiting esterolytic and proteolytic activities of trypsin, by forming tight and stable complexes with the protease through non-covalent interactions.

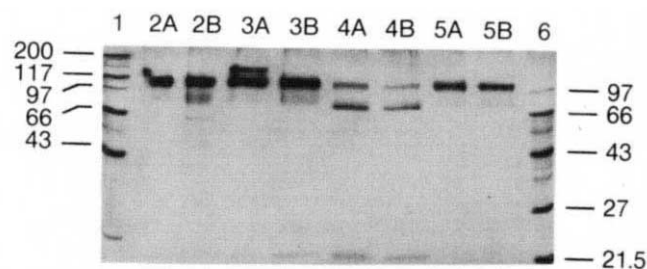


FIG. 4 SDS-PAGE analysis of trypsin and secreted APP mixtures. Lanes 1 and 6, Bio-Rad high protein standards (as described in Fig. 2) and low range protein standards (phosphorylase b, 97 K; bovine serum albumin, 66 K; ovalbumin, 43 K; carbonic anhydrase, 27 K; and soybean trypsin inhibitor, 21.5 K), respectively. Lanes 2A and B, Secreted APP751. Lanes 3A and B, Secreted APP751 and trypsin. Lanes 4A and B, secreted APP695 and trypsin. Lanes 5A and B, secreted APP695. All 'A' lanes indicate samples taken up in Laemmli buffer and not heated, and 'B' lanes refer to identical samples heated to 60 °C before electrophoresis. Secreted APP (150 ng) was mixed with either buffer (20 mM HEPES, pH 7.25) or bovine pancreatic trypsin (40 ng) in duplicate and on ice for 2 min. Tosyl-l-lysyl-chloromethylketone was then added to 1 mM to block any uncomplexed trypsin. Mixtures were taken up in non-reducing Laemmli buffer, and either heated at 60 °C for 5 min or left at room temperature before electrophoresis. The samples were run on 12% acrylamide SDS-PAGE in a Bio-Rad Mini-Protean apparatus and then silver-stained²⁹. In Lane 3A a high relative molecular mass complex of trypsin with secreted APP751 is visible, migrating with a lower mobility than uncomplexed secreted APP751. Under denaturing conditions, however, this complex is not detectable and a free trypsin band is visible at a position corresponding to a M_r of 22 K (lane 3B). With secreted APP695 no complex is detectable and trypsin migrates at its expected position, whether the sample had been previously heated (lane 4B) or not (lane 4A). Comparison with lanes 5A and B shows that the brief incubation with active trypsin in fact led to rapid fragmentation of this secreted APP. An inhibitory complex is formed with secreted APP751, because APP751 is unaffected by its incubation with trypsin.

The sequence identity, similar relative molecular masses, quantitative binding to, and elution with increasing ionic strength from the dextran sulphate matrix, and stable inhibitory complex formation with trypsin, all indicate that the secreted version of APP containing the Kunitz inhibitor domain is indeed PN-II. At the moment we cannot determine if the described PN-II from human fibroblasts⁸ is APP751 or APP770; both of these APPs contain the Kunitz domain and differ only by 19 amino acids. It is probable that material purified from fibroblasts contains a mixture of the two, because all cell types examined to date by S1 nuclease protection assays make both APP751 and APP770, and both of the proteins are poorly resolved by SDS-PAGE²³.

The dissociation of the secreted APP751-trypsin complex under denaturing conditions is not, at first sight, in accord with earlier reports^{8,14}. These described PN-II as a 'serpin' inhibitor (see ref. 24 for review) that forms covalent bonds with target proteases stable to denaturing SDS-PAGE, as does protease nexin I (PN-I), for example²⁵. This classification was based, however, entirely on the detection of an 'SDS-stable' complex of protease with PN-II electrophoresed after mild denaturation^{8,14}. Our results indicate that although complexes of trypsin with PN-II may seem to be SDS-stable if mild denaturing conditions are used, they do not resist heating in 1% SDS, which is the accepted criterion for serpin-protease complex stability²⁶. There is, therefore, no convincing evidence that the original classification of PN-II as a serpin⁸ is correct.

It is ironic that a functional relationship between nexins and APPs containing Kunitz domains has been speculated previously²⁷. Previous comparisons, however, were made between APP and PN-I, a cell-secreted protease inhibitor with neurite outgrowth promoting activity¹³, but which does not have any significant sequence homology with PN-II. The identification of a secreted APP as PN-II shows that Kunitz inhibitors may also be used by cells as 'nexins', and that using PN-I as a mechanistic prototype for uncharacterized potential nexins may be misleading.

The identification of a secreted form of APP as a protease nexin raises new questions regarding the pathogenesis of Alzheimer's disease. It indicates that not only may the protease inhibitory activity of APP be important, but that the reuptake and degradation of APP-protease complexes, characteristic of PN-I, may also be of consequence. In addition, the possible role of PN-II on the uptake of growth-factor-associated serine esterases may have implications for Alzheimer's disease, where changes in such factors could influence the degree of cell survival, neuritic outgrowth, and other cellular processes. Identification and characterization of such variables could lead to a better understanding of the disease and ultimately to its therapy. □

Received 7 June; accepted 31 July 1989.

- Glenner, G. G. & Wong, C. W. *Biochem. biophys. Res. Commun.* **120**, 885-890 (1984).
- Kang, J. *et al. Nature* **325**, 733-736 (1987).
- Tanzi, R. E. *et al. Science* **235**, 880-884 (1987).
- Goldgaber, D., Lerman, M. I., McBride, O. W., Saffiotti, U. & Gajdusek, D. C. *Science* **235**, 877-880 (1987).
- Tanzi, R. E. *et al. Nature* **331**, 528-530 (1988).
- Ponte, P. *et al. Nature* **331**, 525-527 (1988).
- Kitaguchi, N., Takahashi, Y., Tokushima, Y., Shiojiri, S. & Ito, H. *Nature* **331**, 530-532 (1988).
- Van Nostrand, W. E. & Cunningham, D. D. *J. Biol. Chem.* **263**, 8508-8514 (1988).
- Schubert, D., Schroeder, R., LaCortiere, M., Saitoh, T. & Cole, G. *Science* **241**, 223-226 (1988).
- Dykes, T. *et al. EMBO J.* **7**, 949-957 (1988).
- Knauer, D. J., Thompson, J. A. & Cunningham, D. D. *J. Cell Physiol.* **117**, 385-396 (1983).
- Knauer, D. J. & Cunningham, D. D. *Trends biochem. Sci.* **9**, 231-233 (1984).
- Monard, D. *Trends Neurol. Sci.* **11**, 541-544 (1988).
- Knauer, D. J. & Cunningham, D. D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2310-2314 (1982).
- Drinkwater, C. C., Evans, B. A. & Richards, R. I. *Biochemistry* **26**, 6750-6756 (1987).
- Isackson, P. J. *et al. Biochemistry* **26**, 2082-2085 (1987).
- Blaber, M., Isackson, P. J. & Bradshaw, R. A. *Biochemistry* **26**, 6742-6749 (1987).
- Sekoe, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 7341-7345 (1988).
- Trowbridge, I. S., Ralph, P. & Bevan, M. J. *Proc. natn. Acad. Sci. U.S.A.* **72**, 157-161 (1975).
- Johnson, D. A. & Travis, J. *Analyt. Biochem.* **72**, 573-576 (1976).
- Laemmli, U. K. *Nature* **227**, 680-685 (1985).
- Moos, M., Jr, Nguyen, N. Y. & Liu, T.-Y. *J. Biol. Chem.* **263**, 6005-6008 (1988).
- Weidemann, A. *et al. Cell* **57**, 115-126 (1989).

- Travis, J. & Salvesen, G. S. A. *Rev. Biochem.* **52**, 655-709 (1983).
- Baker, J. B., Low, D. A., Simmer, R. L. & Cunningham, D. D. *Cell* **21**, 37-45 (1980).
- Travis, J., Garner, D. & Bowen, J. *Biochemistry* **17**, 5647-5651 (1978).
- Carrell, R. W. *Nature* **331**, 478-479 (1988).
- Seedorf, K., Oltersdorf, T., Krammer, G. & Rowekamp, W. *EMBO J.* **6**, 139-144 (1987).
- Morrissey, J. H. *Analyt. Biochem.* **117**, 307-310 (1981).

ACKNOWLEDGEMENTS. We thank Dr Rachel L. Neve (Children's Hospital, Boston) for APP cDNAs, Dr H.-U. Bernard (National University of Singapore) for the eukaryotic expression plasmid pORFex, Dr W. Rowekamp (German Cancer Research Centre, Heidelberg) for the prokaryotic expression plasmid pex, and Dr Dennis Selkoe (Brigham & Women's Hospital, Boston) for helpful discussions. We especially thank Dr James Travis (University of Georgia, Athens) for reading the manuscript. This work was supported by Athena Neurosciences, Inc., and Eli Lilly & Co.

Developmental alterations in the frequency map of the mammalian cochlea

Stephen M. Echter, Ellis Arjmand & Peter Dallos

Auditory Physiology Laboratory, Department of Neurobiology and Physiology, Northwestern University, Evanston, Illinois 60208, USA

THE position of an auditory hair cell along the length of the cochlea determines the sound frequency to which it is most sensitive. Receptors located near the proximal end (base) of the cochlea are maximally stimulated by high-frequency sounds; those occupying successively more distal (apical) positions respond best to progressively lower frequencies¹. At present, it is unclear how this frequency place map emerges with respect to the development of the cochlea. It has been suggested, on the basis of acoustic trauma experiments with developing chicks² and cochlear potential recordings³⁻⁵ from developing gerbils, that this map may arise through systematic changes in the spatial encoding of frequency along the cochlea. Others have inferred from frequency tuning curves derived from auditory-nerve recordings in developing mammals⁶ and chicks⁷, that the cochlear frequency-place map remains stable throughout development. We analysed frequency tuning curves obtained from gerbil spiral ganglion cells at a constant location within the basal cochlea, and report here that these cells undergo significant increases (up to 1.5 octaves) in their best-response frequencies between the second and third weeks of postnatal life. These recordings provide direct evidence for developmental changes in the tonotopic organization of the mammalian cochlea.

The Mongolian gerbil (*Meriones unguiculatus*) is useful for studies of auditory development because it is of a species in which hearing matures after birth. The first sound-evoked cochlear potentials are recorded on postnatal day 12 (ref. 8), by which time the cochlea is already its mature length (S. Fikret and J. Siegel, unpublished results; D. Harris *et al.*, manuscript in preparation). Two days later, action potentials can be routinely recorded in the auditory nerve⁹. In previous experiments, cochlear potential recordings obtained from the mid-basal turn in an age-graded series of gerbils have shown that the best-response frequency of this recording site changes systematically from ~5 kHz at 13 days after birth to ~17 kHz at 30 days after birth³⁻⁵. Similar alterations in the tonotopic organization of several gerbil auditory brainstem nuclei have also been noted during this time period^{10,11}. Because cochlear length is fully developed and the middle ear is essentially mature by postnatal day 14 (ref. 12), these data imply some sort of reorganization of the high-frequency responsive region of the mammalian cochlea during development. To directly test this hypothesis, we recorded frequency tuning curves from single spiral ganglion cells of known location within the cochlea to assess their best-response frequencies at different stages of development.

Experiments were on gerbils ranging in age from 14 to 53 days. Figure 1 shows the recording approach to Rosenthal's canal within the basal turn of the gerbil cochlea. The recording

site, illustrated schematically, was verified in each case by extracellular injection of horseradish peroxidase. At all ages studied, the peripheral neurites of labelled spiral ganglion cells were orientated radially within the cochlear spiral and projected to inner hair cells located 2.7 ± 0.2 mm from the basal end of the basilar membrane. Total basilar membrane length averages 12.1 mm in this species¹³.

The postnatal development of frequency tuning for single spiral ganglion cells at this recording site is illustrated in Fig. 2. At 14 days after birth, frequency tuning (neural isoresponse) curves possessed a broadly-tuned low-frequency 'tail' and a more sharply-tuned high-frequency 'tip' centred between 5 kHz and 7 kHz (Fig. 2a). Neural thresholds at this age exceeded 90 dB. Over the next three postnatal days, the tip of the tuning curves obtained from neurons at this location shifted progressively upwards in frequency and downwards in threshold (Fig. 2b, c). Tuning curve tails, however, remained relatively stable in their morphology throughout this period, undergoing only a 20-dB reduction in threshold. Neural tuning had stabilized at this location by postnatal day 17. To emphasize this point

we have superimposed the tuning curves derived from 17-, 22- and 53-day old animals (Fig. 2c). This also illustrates both the consistency of our recording site and the homogeneity in tuning-curve morphology observed in adult animals.

In Fig. 3a the characteristic frequencies (CF, the frequency of the lowest threshold) of individual spiral ganglion neurons are plotted as a function of postnatal age. Within a three-day period, from postnatal day 14 to 17, cochlear neurons at this recording location within the basal turn undergo a substantial increase (1.5 octaves) in their CF. The rate at which the CF matures, however, is not constant throughout this postnatal period. The most dramatic increase occurred within a 24-hour period between 14 and 15 days after birth and amounted to ~ 1.0 octave. Over the next two days, the best-response frequency for neurons at this cochlear location continued to increase by ~ 0.5 octave until mature CFs were achieved on postnatal day 17. During this time, neural thresholds at CF decreased by over 80 dB (Fig. 3b).

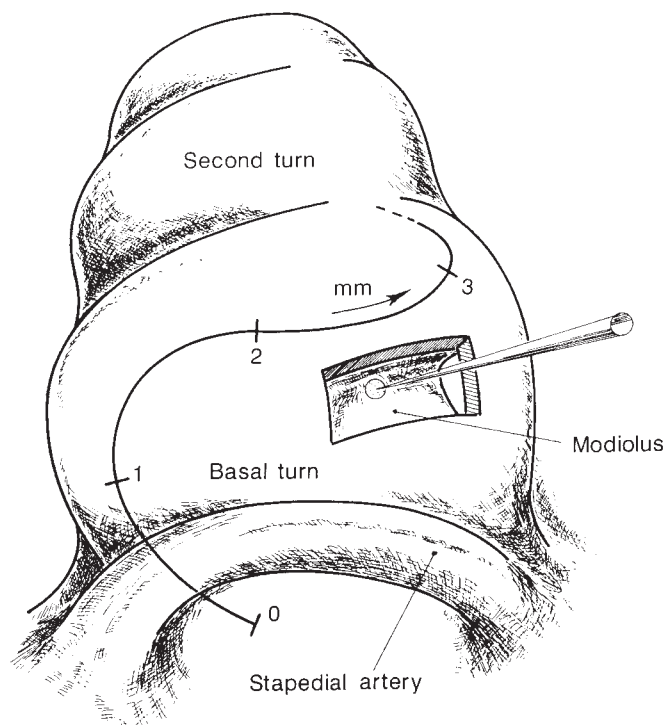


FIG. 1 Schematic illustration of the spiral ganglion recording site within the basal turn of the gerbil cochlea. The basilar membrane is depicted as a spiraling solid line with distance from its base indicated in millimetres. A microelectrode enters scala tympani through a window in the cochlear capsule and passes through an opening in the bone overlying Rosenthal's canal.

METHODS. Gerbils were anaesthetized by injection of sodium pentobarbital (48 mg kg^{-1} , intraperitoneal) and, after the insertion of a tracheal cannula, the right pinna was removed to allow direct coupling of the sound delivery system to the external meatus. The cochlea was exposed using surgical procedures previously described¹⁸. A fine scalpel blade was used to open a rectangular window ($\sim 0.8 \text{ mm} \times 0.5 \text{ mm}$) in the cochlear capsule overlying scala tympani of the basal turn. This window was positioned $\sim 180^\circ$ apical to the round window and its removal allowed visualization of an $800\text{-}\mu\text{m}$ segment of the modiolus within the middle third of the basal turn. Rosenthal's canal, which contains the cell bodies of the spiral ganglion neurons, was easily distinguished as a pearly-grey stripe, $\sim 250\text{-}\mu\text{m}$ wide, coursing through the white modiolar bone. Using visual landmarks, such as modiolar blood vessels, a $100\text{-}\mu\text{m}$ hole was made in the bone overlying a specific portion of Rosenthal's canal using a sharpened insect pin mounted on a three-stage micromanipulator. After surgery, the animal was placed on a vibration-damped table within a double-walled sound-isolation chamber.

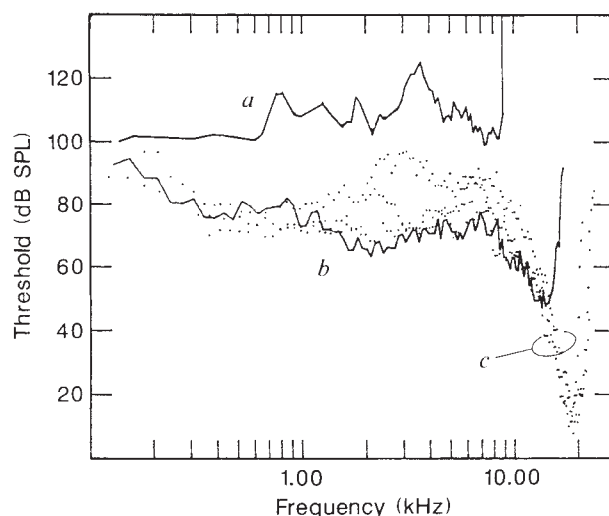


FIG. 2 Tuning curves obtained from single spiral ganglion neurons at the recording location described in Fig. 1, 14 days after birth (a), 15 days after birth (b), and 17, 22, and 53 days after birth (c).

METHODS. Acoustic stimuli were delivered through a Beyer DT48 earphone sealed to the external meatus. Stimulus sound pressure level was monitored at the mouth of this tube, near the tympanum, through a concentrically placed calibrated probe tube coupled to a Knowles microphone. Auditory stimuli were tone-bursts (100-ms duration; 2.5-ms rise/fall time) ranging in frequency from 100 Hz to 21 kHz and delivered at a rate of 3 per second. Stimulus frequency and intensity were controlled by a PDP 11/34 digital computer. Grason-Stadler logic elements controlled stimulus shaping and presentation rate. Single unit activity was recorded with glass microelectrodes filled with NaCl (2M) (impedances of 40–90 M Ω). Neural discharges were amplified and then shaped by a Schmitt trigger, which provided the input to a microprocessor-based event-time counter interfaced to the computer. An automated paradigm was used to derive frequency tuning curves. In this paradigm, the number of spikes was compared in two sequential 50-ms time-windows. A tone was present only during the first window. Threshold was defined as the stimulus intensity required to elicit at least two more spikes in the first window than the second. Stimulus intensity was reduced or increased in 2-dB steps until the threshold criterion was met once. Threshold was determined at up to 135 discrete frequencies per tuning curve. After neural recording, a solution of horseradish peroxidase (Sigma VI, 10% w/v) was iontophoretically ejected ($+2\text{ }\mu\text{A}$ for 15 min; 200-ms pulses, 50% duty cycle) from a glass micropipette ($5\text{-}\mu\text{m}$ tip) to mark the recording location within the spiral ganglion. One to three hours after horseradish peroxidase injection, the animals were killed by overdose of anaesthetic and perfused through the heart with 0.9% saline followed by 3% glutaraldehyde in 0.1 M phosphate buffer. The horseradish peroxidase histochemistry of the cochlea was observed by the reaction with diaminobenzidine with cobalt chloride intensification¹⁹. Cochleae were embedded in Epon and prepared for viewing as a surface preparation. The position of the recording location within the cochlear spiral was computed from tracings made of the basal turn.

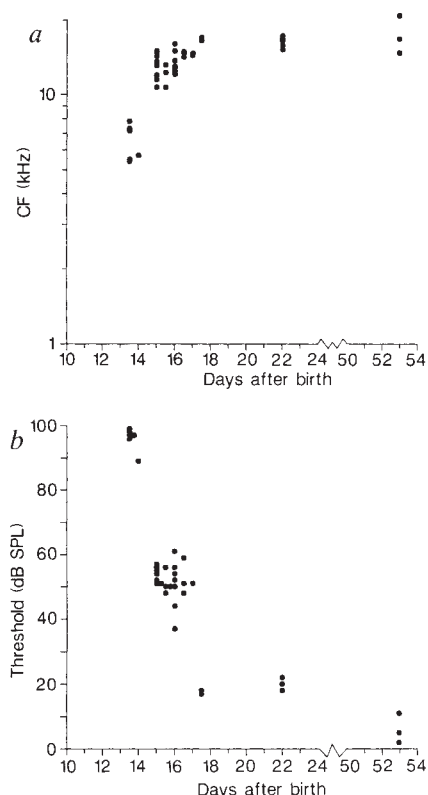


FIG. 3 The characteristic frequencies (CF) of individual spiral ganglion cells (a) and their neural thresholds at CF (b) as a function of postnatal age at the recording site described in Fig. 1. Breeding colonies were checked twice daily. Age was determined with an accuracy of ± 12 hours.

These results confirm the hypothesis that mature tonotopy develops within the inner ear through systematic changes in the spatial encoding of frequency along the cochlear duct². As noted earlier, the first evidence for ontogenic changes in the frequency distribution of the mammalian ear was provided by cochlear potential measurements in developing gerbils³⁻⁵. The relevance of those studies to the emergence of cochlear nerve responses has been questioned¹⁴, however, because they relied on measurements of cochlear microphonic and summing potentials, which are generated primarily by outer hair-cell receptors¹⁵, whereas auditory nerve recordings reflect inner hair-cell responses¹⁶. The results reported here provide direct evidence for developmental changes in tonotopic organization within the auditory nerve. Moreover, the magnitude and time course of the changes in frequency tuning observed in our spiral-ganglion recordings are in substantial agreement with those obtained from cochlear potential measurements, thereby providing additional support for the value of those techniques in mapping functional changes within the developing auditory periphery.

The extent to which our results can be generalized to other mammalian species remains to be determined. Our present knowledge of the functional development of the cochlear nerve in mammals has been derived from extracellular recordings from single or multiple nerve fibres within the auditory nerve trunk. Such techniques, unlike those that we used, cannot verify the exact site within the cochlear spiral from which the neurons recorded from originate. Nevertheless, the emergence of response properties for particular frequency bands has been inferred from many of these studies by comparing tuning curves with the same apparent CF at different developmental ages^{6,17}; this inference presumed a stable cochlear frequency map. If alterations in the frequency-place map such as those reported here are a common feature during cochlear development in mammals, then such studies will require re-evaluation. □

Received 30 May; accepted 24 July 1989.

1. Von Békésy, G. & Rosenblith, W. in *Handbook of Experimental Psychology* (ed. Stevens, J. S.) 1075-1115 (Wiley, New York, 1951).
2. Rubel, E. W. & Ryals, B. M. *Science* **219**, 512-514 (1983).
3. Harris, D. M. & Dallos, P. *Science* **225**, 741-743 (1984).
4. Yancey, C. & Dallos, P. *Hearing Res.* **18**, 189-195 (1985).
5. Arjmand, E., Harris, D. & Dallos, P. *Hearing Res.* **32**, 93-96 (1988).
6. Romand, R. *Neurosci. Lett.* **35**, 271-276 (1983).
7. Manley, G. A., Brix, J. & Kaiser, A. *Science* **237**, 655-656 (1987).
8. Woolf, N. & Ryan, A. *Hearing Res.* **13**, 277-283 (1984).
9. Woolf, N. & Ryan, A. *Devl Brain Res.* **17**, 131-147 (1985).
10. Sanes, D. H. & Rubel, E. W. *J. comp. Neurol.* **279**, 436-444 (1989).
11. Ryan, A. F. & Woolf, N. K. *Devl Brain Res.* **41**, 61-70 (1988).
12. Woolf, N. K. & Ryan, A. F. *Hearing Res.* **35**, 131-142 (1988).
13. Schmiedt, R. A. & Zwislocki, J. J. *J. acoust. Soc. Am.* **61**, 133-149 (1977).
14. Romand, R. *Hearing Res.* **28**, 117-123 (1987).
15. Dallos, P. & Cheatham, M. A. *J. acoust. Soc. Am.* **60**, 510-512 (1976).
16. Liberman, M. C. *Science* **216**, 1239-1241 (1982).
17. Dolan, D. F., Teas, D. C. & Walton, J. P. *J. acoust. Soc. Am.* **78**, 544-554 (1985).
18. Chamberlain, S. C. *J. comp. Neurol.* **171**, 193-204 (1977).
19. Adams, J. J. *Histochem. Cytochem.* **29**, 775 (1981).

ACKNOWLEDGEMENT. This work was supported by a grant from the NIH to P.D.

Molecular cloning and expression of brain-derived neurotrophic factor

Joachim Leibrock, Friedrich Lottspeich*,
Andreas Hohn, Magdalena Hofer,
Bastian Hengerer, Piotr Masiakowski†,
Hans Thoenen & Yves-Alain Barde‡

Max-Planck Institute for Psychiatry, Department of Neurochemistry, and
*Max-Planck Institute for Biochemistry, GeneCenter,
Am Klopferspitz 18A, 8033 Planegg-Martinsried, FRG

DURING the development of the vertebrate nervous system, many neurons depend for survival on interactions with their target cells¹. Specific proteins are thought to be released by the target cells and to play an essential role in these interactions. So far, only one such protein, nerve growth factor, has been fully characterized. This has been possible because of the extraordinarily (and unexplained) large quantities of this protein in some adult tissues that are of no relevance to the developing nervous system². Whereas the dependency of many neurons on their target cells for normal development, and the restricted neuronal specificity of nerve growth factor have long suggested the existence of other such proteins, their low abundance has rendered their characterization difficult. Here we report the full primary structure of brain-derived neurotrophic factor. This very rare protein is known to promote the survival of neuronal populations that are all located either in the central nervous system or directly connected with it³. The messenger RNA for brain-derived neurotrophic factor was found predominantly in the central nervous system, and the sequence of the protein indicates that it is structurally related to nerve growth factor. These results establish that these two neurotrophic factors are related both functionally and structurally.

Brain-derived neurotrophic factor (BDNF) was purified from pig brain^{4,5}, and a partial amino-acid sequence determined both from the N-terminal end and from fragments purified after cleavages (Table 1). The longest sequence, compiled from several overlapping fragments, was used to synthesize two sets of oligonucleotides that were used to prime the amplification of a pig genomic template using the polymerase chain reaction (PCR)⁶. The nucleotide sequence between the two primers was determined and used to synthesize specific primers for further PCRs on a complementary DNA template obtained by reverse

† Present address: Regeneron Pharmaceuticals Inc., 777 Old Saw Mill River Road, Tarrytown, New York 10591-6707, USA.

‡ To whom correspondence should be addressed.

FIG. 1 Nucleotide sequence and deduced amino-acid sequence of BDNF. The peptide sequences obtained from microsequencing (Table 1) are underlined. The only consensus sequence for N-glycosylation is doubly underlined and an arrow indicates the start of mature BDNF.

METHODS. Pig genomic DNA was isolated according to Herrmann and Frischau²². Two separate mixtures of 17-mer oligonucleotides (corresponding to all possible codons and designated oligo 1 and oligo 2) based on the sequence Ala-Ala-Asp-Lys-Lys-Thr (sense) and Lys-Gln-Tyr-Phe-Tyr-Glu (antisense) were chemically synthesized, and added to 1 µg pig genomic DNA template. Amplification was performed using the PCR according to the manufacturer's instructions (GeneAmpTM, Perkin-Elmer Cetus). A DNA band of the predicted size (101 base pairs) was obtained and asymmetrically amplified as described by Innis²³. The resulting sense and antisense DNA fragments were sequenced by the dideoxynucleotide chain termination method²⁴ using the ³²P-end-labelled oligonucleotides 1 and 2 as primers. The nucleotide sequences thus obtained were used to synthesize a sense and an antisense primer to obtain 3'- and 5'- cDNA clones according to the protocol described by Frohman²⁵. The cDNA template was prepared using total RNA extracted (see Fig. 3, legend) from pig brain superior colliculus. Total RNA (80 µg) was transcribed with the reverse transcriptase of the Moloney murine leukaemia virus (BRL, according to the manufacturer's instructions, except for the addition of 1 µl RNase inhibitor (RNasin, Promega) and the omission of actinomycin D), using the primer mixture CGGATCCGAATCTGCAGTTTTTTTTTTT with A, C or G as the terminal 3' nucleotide (oligo 3, designed to match a 3' poly(A) stretch and containing recognition sites for *Bam*HI, *Eco*RI and *Pst*I). The 3' amplified DNA was obtained by using 10 µl of the reverse transcription reaction, the sense primer AACTAGTCGACGGCAGTGGACATGTCCGG (oligo 4, including recognition sites for *Spe*I and *Sal*I) and oligo 3 as the antisense primer. The product giving a hybridization signal with the ³²P-end-labelled oligonucleotide AAG-GATCCTGCAGTTGGCTTTTCGAGACGG (oligo 5, used as an antisense primer in the 5' reaction described below and containing recognition sequences for *Bam*HI and *Pst*I) was digested with *Eco*RI and *Sal*I, cloned into a pBluescript SK⁺ vector (Stratagene) and sequenced. The 5' sequence was obtained using a cDNA template prepared with 80 µg RNA and oligo 2 as the antisense

	AAACCGGG	8
CACCAAGATATCCCTTACCCCTTCTTTTGACCAAGGGAACGTGAAAAAATATAGAGTCTGGGATTCGGGGCC		84
GAAGTCTCCCGAGAGCAGTGCCTTGATGTTTACCTTGACAGTAGTACTGAAAGTTCACAGGTGAGAGAGAGT		166
Met Thr Ile Leu Phe Leu Thr Met Val Ile Ser Tyr Phe Gly Cys Met Lys Ala Pro		20
ATG ACC ATC CTT TTC CTT ACT ATG GTT ATT TCA TCA TCT GGT TGC ATG AAG GCT CCC		226
Met Lys Glu Ala Asn Val Arg Gly Gln Gly Ser Leu Ala Tyr Pro Gly Val Arg Thr His		40
ATG AAA GAA GCC AAC GTC CGA GGA CAA GGC AGC TTG GCC TAC CCA GGT GTG CGG ACC CAT		286
Gly Thr Leu Glu Ser Val Asn Gly Pro Lys Ala Gly Ser Arg Gly Leu Thr Ser Ser Ser		60
GGG ACT CTG GAG AGC GTG AAT GGG CCC AAG GCA GGT TCA AGA GGC CTG ACA TCG TCG TCA		346
Ser Ser Ser Leu Ala Asp Thr Phe Glu His Val Ile Glu Glu Leu Leu Asp Glu Asp Gln		80
TCG TCG TCG TTG GCG GAC ACT TTT GAA CAC GTG ATC GAG GAG CTG TTG GAC GAG GAC CAG		406
Lys Val Arg Pro Asn Glu Glu Asn Asn Lys Asp Ala Asp Met Tyr Thr Ser Arg Val Met		90
AAA GTT CGG CCC AAT GAG GAA AAC AAT AAG GAC GCG GAC ATG TAT ACG TCC CGA GTC ATG		466
Leu Ser Ser Gln Val Pro Leu Glu Pro Pro Leu Leu Phe Leu Leu Glu Glu Tyr Lys Asn		120
CTC AGC AAT GAA GTG CCT TTG GAG CCT CCT CTG CTC TTT CTG CTG GAG GAA TAC AAA AAT		526
Tyr Leu Asp Ala Ala Asn Met Ser Met Arg Val Arg Arg His Ser Asp Pro Ala Arg Arg		140
TAC CTG GAT GCT GCA AAC ATG TCC ATG AGG CTG CCG GCG CAC TCG GAC CCC GCC CCG CGC		586
Gly Glu Leu Ser Val Cys Asp Ser Ile Ser Glu Trp Val Thr Ala Ala Asp Lys Lys Thr		160
GGG GAG CTG AGC ATG TGC GAC AGC ATT AGC GAG TGG GTG ACG GCG GCG GAT AAA GAG AGC		646
Ala Val Asp Met Ser Gly Gly Thr Val Thr Val Leu Glu Lys Val Pro Val Ser Lys Gly		180
GCA GTG GAC ATG CTG GGT GGC ACG GTC ACG GTC CTC GAA AAA GTC CCC GTC TCG AAA GGC		706
Gln Leu Lys Gln Tyr Phe Tyr Glu Thr Lys Cys Asn Pro Met Gly Tyr Thr Lys Glu Gly		200
CAA CTG AAG CAG TAC TTC TAC GAG ACC AAG GAG TGC TGC GCG GAT ACA AAG GAG GGC		766
Cys Arg Gly Ile Asp Lys Arg His Trp Asn Ser Gln Cys Arg Thr Thr Gln Ser Tyr Val		220
TGC AGG GGC ATA GAC AAG AGG CAC TGG AAC TCC CAG TGC CGA ACT ACC CAG TCG TAT GTG		826
Arg Ala Leu Thr Met Asp Ser Lys Lys Arg Ile Gly Trp Arg Phe Ile Arg Ile Arg Met		240
CGG GCC CTC ACC ATG GAT AGC AAA AAA CGA ATT GCG TGG CCG TTC ATA AGG ATA GAC ACT		886
Ser Cys Val Cys Thr Leu Thr Ile Lys Arg Gly Arg End		260
TCC TGT GTA TGT ACT TTG ACC ATT AAG AGG GGA AGA TAG TGGCTTATGTGTATAGATATATTTG		954
AGACAAAAATATCTATTGTATATATACATAACAGGGTAATATTCAGTTAAGAAAAAATATTTATGAACATG		1035
ATGTATAAATGAAGTTTATACAGTACAGTGGTCTACATACTATTTATTTGGACATTTCATGACGAGGGAACAGTC		1115
ATTTTGTGGCACAACCTTAAAAAAGTCTGCATTACCTTCGATAATGTTGGTGTGTTGGCTGCT		1184

primer. An amplification reaction was performed using this template with oligo 3 and oligo 5 as primers and cloned into a pBluescript vector. The inserts of several 3'- and 5'- clones were sequenced.

FIG. 2 Sequence comparison between NGF and BDNF. Areas with more than two amino acids found at identical positions are indicated by shading. The sequences start with the first amino acids of the mature proteins and end with the last ones before the stop codons. Fifty-one amino acids are common to BDNF and the various NGFs¹¹, including the six cysteine residues (see text for details). Asterisks indicate gaps introduced to optimize matching.

HSDPARRGELSVCDISIEWVTAADKKTAIVMSGGTVTVLEKVPVSKGQLKQYFYETKCNP Pig BDNF
 SSSHPVFHHRGEFVSVDISVSVV**GDKTTATDIKGEVNVLEVNINNSVFKQYFFETKCRD Human NGF
 SSSHPVFHHRGEFVSVDISVSVV**GDKTTATDIKGEVNVLEVNINNSVFKQYFFETKCRD Bovine NGF
 SSSHPVFHHRGEFVSVDISVSVV**ADKTTATDIKGEVTVLAEVNVNNVFKQYFFETKCRD Guinea pig NGF
 SSSHPVFHHRGEFVSVDISVSVV**GDKTTATDIKGEVTVLAEVNVNNVFKQYFFETKCRD Mouse NGF
 TADPVLHHRGEFVSVDISVSVV**GDKTTATDIKGEVTVLAEVNVNNVFKQYFFETKCRD Chick NGF
 EDHPVHNLGEHVSVDISVSAW***TKTTATDIKGNVTVMENVNLDNKVYKQYFFETKCKN Snake NGF

MGYTKGCRGIDKRHNWSQCRTTQSYVRALTMDSKKRIGWRFIRIDTSCVCTLTIKRGR Pig BDNF
 PNPVDSGCRGIDSKHNWSYCTTTHTFVKALTMDEGKQ*AAWRFIRIDTACVCLSRKAVRRA Human NGF
 PNPVDSGCRGIDAKHNWSYCTTTHTFVKALTMDEGKQ*AAWRFIRIDTACVCLSRKTGORA Guinea pig NGF
 PSPVESGCRGIDSKHNWSYCTTTHTFVKALTDNKKQ*AAWRFIRIDTACVCLNRKAARRG Bovine NGF
 SNPVESGCRGIDSKHNWSYCTTTHTFVKALTTDEKQ*AAWRFIRIDTACVCLSRKATRRG Mouse NGF
 PRPVSSGCRGIDAKHNWSYCTTTHTFVKALTMDEGKQ*AAWRFIRIDTACVCLSRKSGRP Chick NGF
 PNPEPSGCRGIDSSHWNWSYCTETDTF IKALTMENQ*ASWRFIRIETACVCLTKKKGN Snake NGF

transcription of total RNA isolated from the superior colliculus of the pig brain. This brain area was used because retinal ganglion cells are known to respond to BDNF⁷; the superior colliculus, one of the major targets of retinal ganglion cells⁸, was therefore assumed to be a site of BDNF synthesis. The nucleotide sequence (resulting from combined cDNA clones) shows an open reading frame coding for a protein of 252 amino acids, starting with the first methionine codon found after four in-frame stop codons (Fig. 1). The N terminus of mature BDNF starts after an Arg-Arg cleavage sequence. Both the predicted size of mature BDNF (relative molecular mass 13,511) as well as its calculated isoelectric point (pI = 9.99) agree well with the results of previous determinations made with BDNF purified by two-dimensional gel electrophoresis⁴. The partial amino-acid sequence of BDNF determined by protein microsequencing (singly underlined in Fig. 1) is in complete agreement with the predicted amino-acid sequence deduced from the cDNA

sequence. To demonstrate that this open reading frame is necessary and sufficient to direct the synthesis of biologically active BDNF, we performed the PCR with oligonucleotide primers corresponding to the 5'- and 3'-most regions of the sequence coding for BDNF on a pig genomic template (Table 2). The resulting amplified product had a size corresponding to that predicted from the cDNA sequence, indicating that no introns are present in this portion of the BDNF gene. The amplified product was cloned into an expression vector and introduced into monkey kidney COS cells. The biological activity secreted by these cells into the culture medium (collected after 48 h) was assayed using spinal sensory neurons from embryonic chick. Quantitation of the neuronal survival in this medium indicated that the biological activity was indistinguishable from that of the purified protein (Table 2). The effects were quantitatively similar to those obtained with the purified protein when used at concentrations giving maximal neuronal survival,

TABLE 1 Experimentally determined peptide sequences of BDNF

Peptide sequences	
N terminus	HSDPARRGELSV
Cleavage by:	
V8	XVTAADKKTAVD (1)
V8	KVPVSKGQLKQFYFYE (2)
CNBr	XGGTIVTVLEKVP(V)(S) (3)
CNBr	GYTKEGXRG (4)
Trypsin	(T)AVDMSGGTVTVLEK (5)
Trypsin	ALTMDSK (6)
Combined sequence	VTAADKKTAVDMSGGTVTVLEKVPVSKGQLKQFYFYE
(1, 2, 3, & 5)	

Amino acids represented by the single-letter code. BDNF was either sequenced directly (two runs with 50 and 70 pmole, as determined by amino-acid analysis, with initial yield of 40 and 50 pmole respectively for the N-terminal histidine) or cleaved (5–10 µg BDNF, from five different preparations) as follows: V8, 1 µg *Staphylococcus aureus* V8 (Miles) was added to 5 µg BDNF in 0.2 M NH_4CO_3 , pH 8.0, containing 10% acetonitrile (total volume, 50 µl) and incubated overnight at room temperature; trypsin, 1 µg tosyl phenylalanine chloromethyl ketone (TPCK)-treated trypsin (bovine pancreas, Sigma Type XIII) was added to 8 µg BDNF in Tris-HCl (0.1 M, pH 8.0) containing 10 mM CaCl_2 (total volume, 40 µl) and incubated overnight at 37 °C; cyanogen bromide (CNBr), 10 µg BDNF were incubated for 3 h at room temperature (total volume, 60 µl) with 10% (w/v) CNBr in 70% (v/v) formic acid (final concentrations). After addition of 500 µl H_2O at the end of the reaction, the sample was concentrated to 50 µl in a speed-vac. Tris-HCl (50 µl, 1.0 M, pH 8.0) were added together with 5 µl β -mercaptoethanol and the sample was incubated overnight at 37 °C. After addition of 5 µl iodomethane, the sample was evaporated to dryness on a speed-vac. Reduction and alkylation of BDNF after CNBr cleavage were found to be necessary: no fragments were obtained without reduction, as shown by HPLC and Swank-Munkres SDS-gel electrophoresis¹⁵. This indicates that disulfide bridges are present in BDNF. After all cleavages, the dried samples were resuspended in 0.1% trifluoroacetic acid (TFA) and applied onto a reverse phase C8 microbore HPLC column (Applied Biosystems) and the peptides eluted at 0.1 ml min⁻¹, using a 60-min linear gradient of 0–60% acetonitrile in 0.1% TFA. Detection was at 214 nm with a Waters 441 ultraviolet detector. The peptides were sequenced by automated Edman degradation on a gas-phase microsequencer (model 470A Applied Biosystems), according to Hewick¹⁶ and Hunkapiller¹⁷. Detection of the phenylthiohydantoin amino acids was as described¹⁸.

TABLE 2 Survival of cultured chick sensory neurons

	COS medium (final dilution)		
	1:20	1:50	1:200
	Cell count		
COS ⁺		2,510 ± 263	2,833 ± 171
COS ⁻	211 ± 16		
COS	250 ± 87		
BDNF + COS ⁺		2,516 ± 209	
NGF + COS ⁺		5,770 ± 72	
BDNF alone		2,718 ± 424	

E8 chick spinal sensory neurons were plated (6,000 per well), incubated for 24 h and counted after 24 h (ref. 9). The values are the means of triplicate determinations ± s.d. BDNF and NGF were used at 1 ng ml⁻¹, concentrations at which maximal neuron survival is observed with either factor. COS⁺ and COS⁻, cells transfected with plasmids containing the BDNF insert in the sense or antisense orientation, respectively. COS, non-transfected cells. At dilutions higher than 1:20, no survival above that of the control was seen with either COS⁻ or COS conditioned medium. In all experiments without NGF, a monoclonal anti-NGF antibody¹⁹ was used at 1 µg ml⁻¹. The sequence corresponding to pig preproBDNF was obtained by using the oligonucleotide primers ATAATCTAGATGACCATCCTTTTCCTT (sense) and ATAATCTAGAC-TATCTTCCCCTCTTAAT (antisense) in a PCR using 1 µg of pig genomic template (each primer had an added *Xba*I site). After digestion with *Xba*I, the amplified DNA was ligated in the *Xba*I site of the plasmid pCMV²⁰ and XL-1 bacteria transfected by electroporation. Plasmid DNA was cut with both *Xba*I and *Pst*II. The size of the resulting products allowed us to determine the orientation of the inserts, and both plasmids were used to transfect COS cells by the calcium phosphate method²¹, and the growth medium (consisting of DMEM and 10% FCS) collected after 24 h.

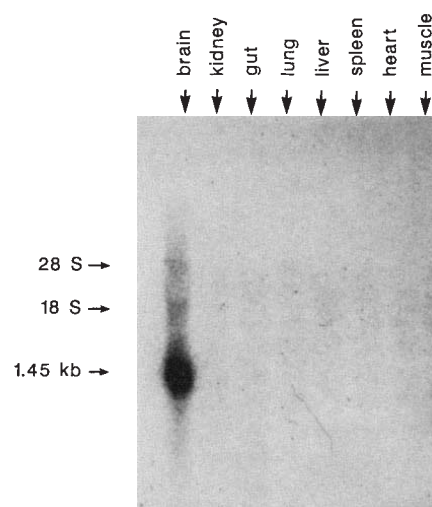


FIG. 3 Northern blot analysis. From each tissue, 20 µg total RNA was applied per lane and hybridized with a ³²P-labelled cRNA mouse BDNF probe. Note that a strong signal can be seen at a position corresponding to ~1.45 kb with brain tissue, but not with the seven other tissues analysed. Muscle is thigh skeletal muscle.

METHODS. Total RNA was extracted from various tissues of adult female mice according to Okayama²⁸. Electrophoresis was according to Lehrach²⁹ on 1.3%-agarose-formaldehyde gels. The RNA was transferred to nylon membranes (Hybond-N, Amersham) and hybridized overnight at 65 °C in 2 ml hybridization buffer (7.5 × SSC, 50% formamide, 5 × Denhardt's solution, 0.25% SDS, 250 µg ml⁻¹ salmon sperm DNA, 5 mM EDTA in 50 mM sodium phosphate, pH 7.2) with a ³²P-cRNA mouse BDNF probe (10⁷ c.p.m. ml⁻¹). Washing was for 60 min at 60 °C in 0.1 × SSC containing 0.5% SDS. After washing, the blot was incubated for 60 min at room temperature with 0.1 µg ml⁻¹ RNase A (Pharmacia) and the film exposed at -70 °C (with intensifying screen) for 48 h. Preparation of the mouse probe: in brief, a mouse brain cDNA library (Clontech) was screened with two independent BDNF oligonucleotides. Double positive clones were isolated and subcloned into the *Eco*RI site of a pBluescript SK⁺ vector (Stratagene). The nucleotide sequence corresponding to nucleotides 350–829 of the pig sequence was determined, and a total of only four amino-acid differences (all in the precursor part) were found between mouse and pig BDNF. This indicates a remarkable degree of sequence conservation between pig and mouse BDNF. A single-stranded RNA probe was prepared using this template and T3 polymerase (Promega). The specific activity of this probe was 10⁸ c.p.m. µg⁻¹.

and were additive to those of NGF, as previously reported for BDNF⁹.

A striking feature of the primary structure of mature BDNF is its similarity to that of NGF: with only three gaps introduced in the NGF sequences to optimize matching, 51 identities are common to the various nerve growth factors (NGFs) (from snake to man) and BDNF (Fig. 2). Significantly, these identities include all six cysteine residues. Although the exact arrangement of the disulfide bridges of BDNF are not yet known, it is clear that such bridges are present (Table 1, legend). Also, the three tryptophan and two phenylalanine residues, as well as most valine and aspartic acid residues, are found at identical positions in NGF.

Most of BDNF's precursor sequence is unrelated to that of NGF, with two exceptions: the putative secretory signal sequence of BDNF shows 5 amino-acid identities (out of 18 amino acids) and a striking overall relatedness with the signal sequence of mouse NGF, where cleavage has been demonstrated to occur after an alanine residue at position 18 (ref. 10). The other similarity with NGF starts at the only N-glycosylation consensus sequence (doubly underlined in Fig. 1), corresponding to asparagine 126. This asparagine is located eight amino acids before the cleavage site giving rise to mature BDNF. The same arrangement is found in several NGFs, as well as the sequence Arg-X-Basic residue-Arg as the last four amino acids of the precursor¹¹.

To study the site of BDNF synthesis, a mouse cRNA probe was used to probe mouse tissues in northern blot analyses. (Rapid tissue processing for RNA extraction, which is technically difficult with pig tissues, was found to be essential.) A strong signal at a position corresponding to 1.45 kilobases (kb) was detected in the brain (Fig. 3) and spinal cord (data not shown), whereas under identical hybridization conditions, no signal was detected in all other tissue tested including kidney, gut, lung, liver, spleen, heart and skeletal muscle (Fig. 3).

Two main conclusions can be drawn from this study. First, the mRNA coding for BDNF is present in the central nervous system, in areas such as the spinal cord and the superior colliculus that contain the target cells of neurons previously shown to respond to BDNF (primary sensory neurones^{9,12} and retinal ganglion cells⁷). These data support the view that BDNF is a target-derived neurotrophic factor. Clearly, detailed studies need to be done to examine the various sites of synthesis of BDNF in the CNS. It is already evident, however, that the distribution of BDNF mRNA is very different from that of NGF mRNA, which is found in many non-CNS tissues^{13,14}. It is also apparent that the amounts of BDNF mRNA in the brain are considerably larger than those of NGF mRNA. Second, the amino-acid sequence of mature BDNF shows striking similarity to that of mature NGF, including all six cysteine residues. It is tempting to speculate that BDNF and NGF might also display similarities in their tertiary structures. Their respective neuronal receptors are thus likely to also have common structural features.

There is no reason to think that BDNF and NGF should be the only members of a family of neurotrophic proteins having in common both functional and structural characteristics. Indeed, from a descriptive point of view, a variety of observations suggest that the control of neuronal survival and maintenance of function exerted by target cells both during development and in adulthood is a general phenomenon, and that this control also applies to NGF- and BDNF-independent neurons. Thus it is hoped that the structural features common to both NGF and BDNF can be used to aid the discovery of other members. □

Regulation of Ca^{2+} -dependent K^{+} -channel activity in tracheal myocytes by phosphorylation

H. Kume, A. Takai[†], H. Tokuno* & T. Tomita*

The Second Department of Internal Medicine and * Department of Physiology, School of Medicine, Nagoya University, Tsurumai-cho 65, Showa-ku, Nagoya 466, Japan

ISOPRENALINE is a β -adrenergic agonist of clinical importance as a remedy for asthma. In airway smooth muscle its relaxant action is accompanied by hyperpolarization of the membrane^{1,2} and elevation of the level of intracellular cyclic AMP³. Hyperpolarization and relaxation are also induced by drugs such as forskolin, theophylline and dibutyryl cAMP, indicating that cAMP-dependent phosphorylation is involved in producing the electrical response¹. Cyclic AMP-dependent protein kinase (protein kinase A) has been reported to activate Ca^{2+} -dependent K^{+} channels in cultured aortic smooth muscle cells⁴ and snail neurons^{5,6}. The membrane of tracheal smooth-muscle cells is characterized by a dense distribution of Ca^{2+} -dependent K^{+} -channels⁷. We have now examined the effect of isoprenaline and protein kinase A on Ca^{2+} -dependent K^{+} -channels in isolated smooth muscle cells of rabbit trachea, using the patch-clamp technique⁸. Our results show that the open-state probability of Ca^{2+} -dependent K^{+} -channel of tracheal myocytes is reversibly increased by either extracellular application of isoprenaline or intracellular application of protein kinase A. We also show that this effect is significantly enhanced and prolonged in the presence of a potent protein phosphatase inhibitor, okadaic acid⁹⁻¹³.

Smooth muscle cells were enzymatically isolated¹⁴ from the trachea of adult male rabbits (2.5–3 kg). The giga-seal method⁸ was applied to measure the membrane current under

[†] To whom correspondence should be addressed.

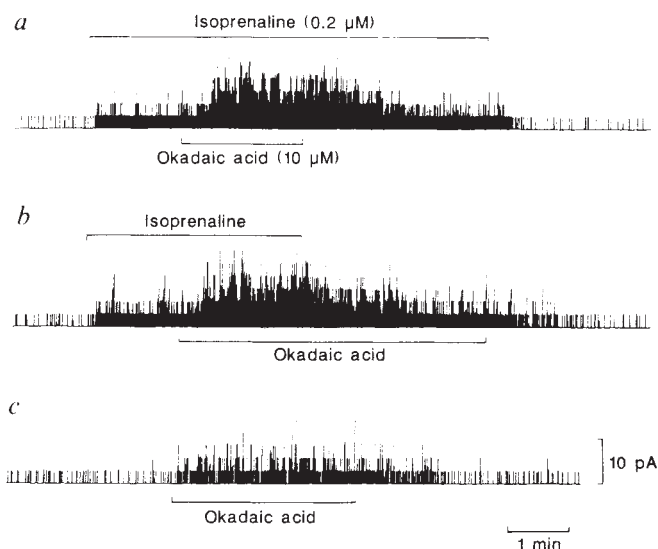


FIG. 1 Effect of isoprenaline on the channel activity and its enhancement by a protein phosphatase inhibitor, okadaic acid. For an explanation of a, b and c, see text. Tracheal smooth muscle cells were dispersed with collagenase and elastase. Continuous recording with a cell-attached pipette filled with a physiological saline of the following composition (mM): NaCl, 127; KCl, 5.9; CaCl_2 , 2.4; MgCl_2 , 1.2; dextrose, 11.8 and HEPES, 10 (pH 7.4 at 25 °C adjusted with NaOH). The isolated myocytes were transferred into a small chamber (0.2 ml) which was perfused with physiological saline at 3 ml min⁻¹. Isoprenaline (0.2 μM) and okadaic acid (10 μM) were added to the perfusing solution.

Received 30 June; accepted 8 August 1989.

- Purves, D. in *Body and Brain. A Tropic Theory of Neural Connections* (Harvard University Press, Cambridge 1988).
- Cohen, S. *Proc. natn. Acad. Sci. U.S.A.* **46**, 302–311 (1960).
- Barde, Y.-A. *Neuron* **2**, 1525–1534 (1989).
- Barde, Y.-A., Edgar, D. & Thoenen, H. *EMBO J.* **1**, 549–553 (1982).
- Hofer, M. M. & Barde, Y.-A. *Nature* **331**, 261–262 (1988).
- Saiki, R. K. *et al. Science* **230**, 1350–1354 (1985).
- Johnson, J. E., Barde, Y.-A., Schwab, M. & Thoenen, H. *J. Neuroscience* **6**, 3031–3038 (1986).
- Campos-Ortega, J. A. *Brain Res.* **19**, 306–312 (1970).
- Lindsay, R. M., Thoenen, H. & Barde, Y.-A. *Dev. Biol.* **112**, 319–328 (1985).
- Edwards, R. H. *et al. Molec. cell. Biol.* **8**, 2456–2464 (1988).
- Schwarz, M. A., Fisher, D., Bradshaw, R. A. & Jackson, P. J. *J. Neurochem.* **52**, 1203–1209 (1989).
- Davies, A. M., Thoenen, H. & Barde, Y.-A. *J. Neurosci.* **6**, 1897–1904 (1986).
- Heumann, R., Korsching, S., Scott, J. & Thoenen, H. *EMBO J.* **3**, 3183–3189 (1984).
- Shelton, D. L. & Reichardt, L. F. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7951–7955 (1984).
- Swank, R. T. & Munkres, K. D. *Analyt. Biochem.* **39**, 462–477 (1971).
- Hewick, R. M., Hunkapiller, M. W., Hood, L. E. & Dreyer, W. J. *J. Biol. Chem.* **256**, 7990–7997 (1981).
- Hunkapiller, M. W. *et al. Meth. Enzym.* **91**, 227–236 (1983).
- Lottspeich, F. *J. Chromat.* **326**, 321–327 (1985).
- Korsching, S. & Thoenen, H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3513–3516 (1983).
- Andersson, S. *et al. J. Biol. Chem.* **264**, 8222–8229 (1989).
- Chen, C. & Okayama, H. *Molec. cell. Biol.* **7**, 2745–2752 (1987).
- Herrmann, G. & Frischauf, A.-M. *Meth. Enzym.* **152**, 180–183 (1987).
- Innis, M. A., Myambo, K. B., Gelfand, D. H. & Brown, M. A. D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9436–9440 (1988).
- Sanger, F., Nickol, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3918–3921 (1979).
- Frohman, M. A., Dush, M. K. & Martin, G. R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8998–9002 (1988).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **137**, 266–267 (1979).
- Scott, J. *et al. Nature* **302**, 538–540 (1983).
- Okayama, H. *et al. Meth. Enzym.* **154**, 3–28 (1987).
- Lehrach, H., Diamond, D., Worner, J. & Boedtker, H. *Biochemistry* **16**, 4743–4751 (1977).

ACKNOWLEDGEMENTS We thank David Edgar for comments on the manuscript, Dan Lindholm and Rolf Heumann for suggestions, Josef Kellermann for amino-acid analyses, and Frau Claudia Cap for technical assistance. This work was supported in part by the Bundesministerium für Forschung und Technologie. Y.A.B. was supported by the Hermann-und-Lilly Schilling Stiftung and P.M. by Regeneron Pharmaceuticals, Inc.

voltage-clamp conditions at room temperature (20–25 °C). Okadaic acid is a polyether derivative of a 38-carbon fatty acid which was first isolated from marine sponges of the genus *Halichondria*¹⁵. It is a potent inhibitor of protein phosphatases 1 and 2A (PP1 and PP2A)^{11,12,16} and produces various effects in physiological systems that are modulated by phosphorylation^{9–11,13,17}. For example, okadaic acid has a strong inotropic action on heart muscle¹⁸ that is due to an increase in the open-state probability of Ca^{2+} channels¹¹, which is thought to be regulated by cAMP-dependent phosphorylation¹⁹. Okadaic acid has no effect on protein kinases, including protein kinase A^{13,17}, protein kinase C²⁰ and Ca^{2+} /calmodulin-dependent myosin light-chain kinase⁹. In smooth muscle tissues, the activities of intracellular PP1 and PP2A are nearly abolished by 10 μM okadaic acid²¹.

Figure 1 shows the effect of extracellular application of isoprenaline and okadaic acid on channel activity recorded with a cell-attached pipette. The voltage in the pipette was clamped at -40 mV . Under this condition, the unitary current through the channel was $\sim 3\text{ pA}$ and the open-state probability (P_0 ; mean $\pm$ s.e.m.) was 0.012 ± 0.004 ($n=3$). Isoprenaline ($0.2\text{ }\mu\text{M}$) increased P_0 to 0.08 ± 0.02 ($n=3$) (Fig. 1a). The P_0 further increased to 0.19 ± 0.08 ($n=3$) when $10\text{ }\mu\text{M}$ okadaic acid was also applied (Fig. 1a). The activating effect of isoprenaline on the channels disappeared within 15–20 s, either by removing isoprenaline (Fig. 1a) or by adding $1\text{ }\mu\text{M}$ propranolol to the perfusing solution (results not shown).

The time course for washing out the effect of isoprenaline was markedly prolonged in the presence of $10\text{ }\mu\text{M}$ okadaic acid

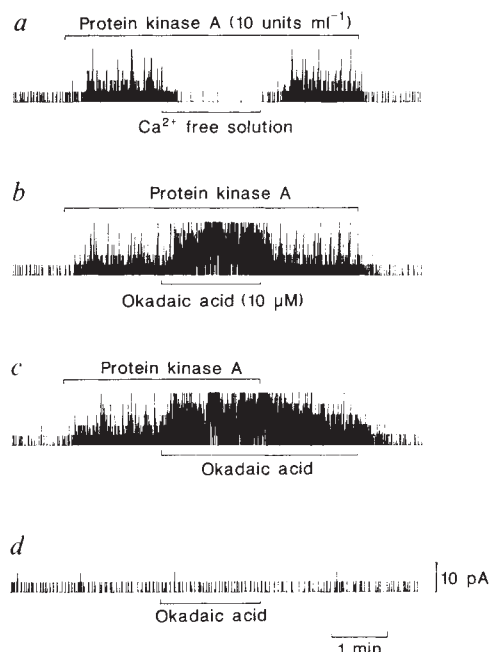


FIG. 2 Effect of protein kinase A on the channel activity and its enhancement by okadaic acid. Continuous recording in an inside-out membrane patch from a tracheal smooth muscle cell at a holding potential of 0 mV . For explanation of a–e, see text. The pipette filled with physiological saline was attached to the cell membrane, and the response to $0.2\text{ }\mu\text{M}$ isoprenaline shown in Fig. 2 was observed before the membrane was isolated from the cell (not shown). After establishment of the inside-out configuration, the cytoplasmic side of the membrane was perfused with a solution containing (mM): potassium aspartate, 80; KCl, 50; NaCl, 20; MgCl_2 , 0.8; HEPES, 10; EGTA, 5; ATP- Na_2 , 0.3; and cAMP, 0.1 (pH 6.9 at $25\text{ }^\circ\text{C}$, adjusted with NaOH). The free Ca^{2+} concentration was adjusted to $0.1\text{ }\mu\text{M}$ by adding CaCl_2 . (For the Ca^{2+} -free solution, CaCl_2 was not added.) Commercial protein kinase A (Sigma P5511; from bovine heart) was dialysed just before use against the perfusing solution at $0\text{ }^\circ\text{C}$ for 4 h to remove buffer materials (such as EDTA, potassium phosphate). Addition of okadaic acid ($10\text{ }\mu\text{M}$) and dialysed protein kinase A (10 units ml^{-1}) had very little effect on the pH of the perfusing solution.

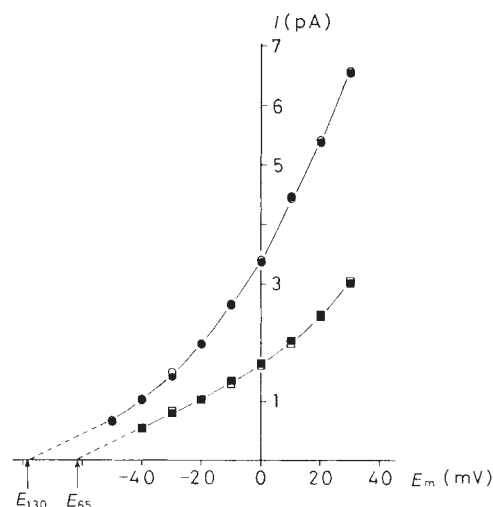


FIG. 3 The I - V relationship of the channel activated by protein kinase A. In the same type of patch-clamp experiment as illustrated in Fig. 2, the amplitude of the unitary current of the channel (I) was measured at various holding potentials (E_m) using the pipette filled with physiological saline ($[\text{K}^+]_o = 5.9\text{ mM}$). The K^+ concentration in the bath solution, $[\text{K}^+]_i$, was reduced from 130 mM ($\circ$ and $\bullet$) to 65 mM ($\square$ and $\blacksquare$) by isosmotically replacing K^+ with Na^+ . Measurements were made in the absence (open symbols) and presence (closed symbols) of protein kinase A (10 units ml^{-1}) on the cytoplasmic side. The equilibrium potential of K^+ ($= (RT/F) \ln([\text{K}^+]_o/[\text{K}^+]_i)$) for $[\text{K}^+]_i$ of 130 mM and 65 mM are denoted by E_{130} and E_{65} , respectively.

(Fig. 1b). The increased P_0 did not return to the control level until okadaic acid was removed 3 min after the removal of isoprenaline (Fig. 1b). The P_0 increased to 0.04 ± 0.01 ($n=3$) when okadaic acid was applied in the absence of isoprenaline (Fig. 1c).

These observations are consistent with the activation of channels by β -adrenergic stimulation that is mediated by phosphorylation of some regulatory protein(s). This possibility was further examined in experiments using isolated inside-out membrane patches (Fig. 2). The cytoplasmic side of the membrane was perfused with a solution containing $0.1\text{ }\mu\text{M}$ free Ca^{2+} and the voltage in the pipette was clamped at 0 mV . The channels were reversibly activated when protein kinase A (10 units ml^{-1}) was applied to the perfusing solution in the presence of 0.1 mM cAMP and 0.3 mM ATP (Fig. 2a). The unitary current of the channels was $3.3 \pm 0.1\text{ pA}$ ($n=10$) and P_0 was 0.008 ± 0.002 ($n=4$). The P_0 increased to 0.033 ± 0.006 ($n=4$) $\sim 20\text{ s}$ after application of protein kinase A. Channel opening was almost completely suppressed by perfusing with Ca^{2+} -free solution in the presence of protein kinase A (Fig. 2a). The P_0 returned to the control level about 10 s after removal of protein kinase A (Fig. 2a), indicating the presence of a relatively high phosphatase activity in the isolated membrane. Indeed, P_0 was increased to 0.23 ± 0.05 ($n=3$) by adding okadaic acid ($10\text{ }\mu\text{M}$), a protein phosphatase inhibitor, to the perfusing solution in the presence of protein kinase A and $0.1\text{ }\mu\text{M}$ free Ca^{2+} (compare Fig. 1). The activating effect of okadaic acid disappeared quickly when okadaic acid was removed (compare Fig. 1a). The time course of disappearance was considerably faster than that observed in Fig. 1a, supporting the idea that the site of action of okadaic acid is intracellular. Okadaic acid significantly slowed the recovery from the activating effect of protein kinase A (Fig. 2c; compare Fig. 1b). In excised patches, okadaic acid ($10\text{ }\mu\text{M}$) failed to produce any effect on channel activity when it was applied in the absence of protein kinase A (Fig. 2d; compare Fig. 1c).

Similar results were obtained when the catalytic subunit of protein kinase A (Sigma P2645; 10 units ml^{-1}) was used instead of the holoenzyme and cAMP. Neither the holoenzyme nor the

catalytic subunit produced any effect on channel activity in the absence of ATP (data not shown).

Figure 3 shows the voltage-current (V - I) relationship for the channel activated by protein kinase A. In an inside-out patch, the unitary current of the channel (I) was measured at various holding potentials (E_m). The concentration of K^+ in the pipette ($[K^+]_o$) was 5.9 mM. When the K^+ concentration in the bath ($[K^+]_i$) was 130 mM, the maximal slope conductance of 120 pS was obtained in the voltage range 20–30 mV (Fig. 3). In myocytes of the rabbit portal vein, two types of Ca^{2+} -dependent K^+ -channels have been identified with either a large or small conductance (K_L type and K_S type)²². The V - I relationship for the channels observed in the present experiments was similar to that for the K_L -type K^+ -channel. The V - I curve shifted to the right and the curvature decreased as $[K^+]_i$ was reduced. The maximal-slope conductance was 55 pS when $[K^+]_i$ was 65 mM (Fig. 3). The reversal potentials estimated for $[K^+]_i$ of 130 mM and 65 mM by extrapolation of the V - I curves to the abscissa, agreed well with the equilibrium potentials of K^+ (E_{130} and E_{65}). Thus, channels activated by protein kinase A observed in these experiments belong to the category of Ca^{2+} -activated K^+ -channels (see also Fig. 2a). The V - I relationship was not affected by protein kinase A (Fig. 3) or okadaic acid.

These results strongly support the idea that isoprenaline activates Ca^{2+} -dependent K^+ -channels in tracheal myocytes by phosphorylation of a channel- or channel-related protein(s). In isolated membranes, okadaic acid increased the open-state probability of the channels in the presence of protein kinase A activity, whereas it failed to produce any effect when protein kinase A was not added to the perfusing solution (see Fig. 2). This indicates that some protein phosphatase activity is bound to the membrane, whereas the kinase activity is present only in the cytosol. In intact myocytes, okadaic acid caused a slight but clear elevation of the P_o in the absence of isoprenaline (see Fig. 1c). Therefore, some kinase activity seems to remain even under conditions of no stimulation.

Until now the clinically important relaxant action of isoprenaline on airway smooth muscle has been thought to be due to a cAMP-dependent increase in sequestration of intracellular Ca^{2+} and/or Ca^{2+} extrusion from the cell^{23,24}. Because an increase of K^+ conductance results in hyperpolarization, the cAMP-mediated activation of the Ca^{2+} -dependent K^+ -channels, which are densely distributed in the plasma membrane of tracheal myocytes⁷, may be another important factor in the relaxant effect of isoprenaline. □

Received 22 May; accepted 26 June 1989

- Honda, K., Satake, T., Takagi, K. & Tomita, T. *Br. J. Pharmacol.* **87**, 665–671 (1986).
- Honda, K. & Tomita, T. *Jap. J. Physiol.* **37**, 333–336 (1987).
- Fredholm, B. B., Brodin, K. & Standberg, K. *Acta pharmacol. tox.* **45**, 336–344 (1979).
- Sadoshima, J., Akaike, N., Kanaide, H. & Nakamura, M. *Am. J. Physiol.* **255**, H754–H759 (1988).
- De Peys, J. E., Cachelin, A. B., Levitan, I. B. & Reuter, H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4207–4211 (1982).
- Ewald, D. A., Williams, A. & Levitan, I. B. *Nature* **315**, 503–506 (1985).
- MacCann, J. & Walsh, M. J. *J. Physiol., Lond.* **372**, 264–270 (1986).
- Hamill, O. P., Marty, A., Neher, E., Sackmann, B. & Sigworth, J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).
- Takai, A., Bialojan, C., Troschka, M. & Rüegg, J. C. *FEBS Lett.* **217**, 81–84 (1987).
- Bialojan, C., Takai, A. & Rüegg, J. C. *J. Physiol., Lond.* **398**, 81–95 (1988).
- Hescheler, J., Mieskes, G., Rüegg, J. C., Takai, A. & Trautwein, W. *Pflügers Arch. ges. Physiol.* **412**, 248–252 (1988).
- Bialojan, C. & Takai, A. *Biochem. J.* **256**, 283–290 (1988).
- Takai, A. *J. Muscle Res. Cell Motil.* **9**, 563–565 (1988).
- Benham, C. D., Bolton, T. B., Lang, R. J. & Takewaki, T. *J. Physiol., Lond.* **371**, 45–67 (1986).
- Tachibana, K. *et al. J. Am. chem. Soc.* **103**, 2469–2471 (1981).
- Ingebritsen, T. S. & Cohen, P. *Eur. J. Biochem.* **132**, 255–261 (1983).
- Haystead, T. A. *et al. Nature* **337**, 78–81 (1989).
- Kodama, I., Kondo, N. & Shibata, S. *J. Physiol., Lond.* **378**, 359–373 (1986).
- Reuter, H. *Nature* **301**, 569–574 (1983).
- Suganuma, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 1768–1771 (1988).
- Takai, A., Troschka, M., Mieskes, G. & Somlyo, A. *Biochem. J.* (in the press).
- Inoue, R., Kitamura, K. & Kuriyama, H. *Pflügers Arch. ges. Physiol.* **405**, 173–179 (1985).
- Hardman, J. G. in *Smooth Muscle: An Assessment of Current Knowledge* (eds Bülbirg, E., Brading, A. F., Jones, A. W. & Tomita, T.) 249–262 (1981).
- Hisayama, T. & Takayanagi, I. *Biochem. Pharmacol.* **32**, 3197–3203 (1983).

ACKNOWLEDGEMENTS. We thank Dr Y. Tsukitani (Fujisawa Pharmaceutical Co., Tokyo) for providing okadaic acid and Dr J. C. Rüegg (Heidelberg) for reading the manuscript. This work was supported by the Ministry of Education of Japan.

Bidirectional movement of actin filaments along tracks of myosin heads

Yoko Yano Toyoshima*, Chikashi Toyoshima* & James A. Spudich†

Department of Cell Biology, Stanford University School of Medicine, Stanford, California 94305, USA

IT is well established that muscle contraction results from the relative sliding of actin and myosin filaments. Both filaments have definite polarities and well-ordered structures. Thick filaments, however, are not vital for supporting movement *in vitro*^{1–4}. Previously we have demonstrated that actin filaments can move continuously on myosin fragments (subfragment-1 or heavy meromyosin (HMM)) that are bound to a nitrocellulose surface³. Here we report that actin filaments can move in opposite directions on tracks of myosin heads formed when actin filaments decorated with HMM are placed on a nitrocellulose surface. The actin filaments always move forward, frequently changing the direction of the movement, but never move backward reversing the polarity of the movement. The direction of movement is therefore determined by the polarity of the actin filament. These results indicate that myosin heads have considerable flexibility.

Decorated actin filaments were formed by mixing HMM and actin filaments labelled with rhodamine-phalloidin at an HMM:actin molar ratio of slightly less than 1:2. The solution was diluted 20 times and applied to a flow cell. A solution of HMM alone, applied at this concentration ($5 \mu\text{g ml}^{-1}$) and randomly distributed over the nitrocellulose surface, was too dilute to support movement (Toyoshima *et al.*, manuscript in preparation). The decorated actin filaments bound to the nitrocellulose surface showed the characteristic arrowhead pattern indicative of a unipolar arrangement of myosin heads along the actin filaments^{5,6} (Fig. 3a).

On addition of ATP, actin filaments started to move. In Fig. 1, initial positions of actin filaments are shown in red, thus indicating the tracks of HMM; moving filaments appear in white (the part of the actin filament on the HMM tracks) or blue (the part that has run off the HMM tracks). After the actin filaments moved one filament length at a constant speed, they came off the surface and showed only Brownian motion (Fig. 1a). This result demonstrates that the inertia of the filament was negligible, as was expected for movement in a low-Reynolds-number fluid. During the following minutes, some of the actin filaments released from HMM tracks elsewhere in the flow cell came to the surface again by diffusion and repeated the movement: they followed exactly the same tracks, and came off at the end of the tracks. Surprisingly, some filaments moved in the opposite direction to that of the initial filaments. Occasionally an actin filament moved to one end of the track and came off, then flipped over and moved towards the other end of the same track. Sometimes two actin filaments passed over each other on a track (Fig. 1b). The speed of the movement was about the same in either direction and independent of the filament length. These observations indicate that the direction of the movement is simply determined by the polarity of actin filaments and that the HMM is flexible enough to support the movement in any direction along the surface.

The density of HMM and the width of the track were examined before and after the addition of ATP by using immunogold labelling. The primary antibody used had a very high affinity for the 25K–50K junction of myosin heads, and the secondary antibody was conjugated with gold. Figure 2a shows a fluores-

* Present addresses: Department of Biology, Ochanomizu University, Ohtsuka, Tokyo, Japan (Y.Y.T.) and Frontier Research Programs, Riken, Wako, Saitama, Japan (C.T.).

† To whom correspondence should be addressed.

FIG. 1 Sequences of video-processed images showing the movement of actin filaments along tracks of HMM. Locations of initial actin filaments decorated with HMM to form the HMM tracks were recorded before ATP addition; they are shown in red and superimposed on real-time images shown in blue. As a result, the part of the filament on the HMM track appears white. *a*, Immediately after the addition of ATP. Filaments moved one filament length and came off the surface (or sometimes remained attached at their trailing ends, possibly due to radiation damage) and exhibited Brownian motion. Sequence, 0.5 s between frames. *b*, Actin filaments following the same track in opposite directions (arrows). In the minutes following addition of ATP, actin filaments released from HMM tracks elsewhere came to the surface again by diffusion and followed the same tracks. In this example, two filaments moved on the track from left to right while one short filament moved from right to left. Sequence, 0.5 s between frames; scale bar, 10 μm .

METHODS. Protein preparations, assay conditions and the optics system have been described previously^{3,14}. Chymotryptic HMM ($100 \mu\text{g ml}^{-1}$) was mixed with rhodamine-phalloidin-labelled actin filaments ($28 \mu\text{g ml}^{-1}$) at slightly less than stoichiometric amounts in the assay buffer (25 mM KCl, 4 mM MgCl_2 , 1 mM dithiothreitol, 25 mM imidazole, pH 7.4) to form rigor complexes. The mixture was applied to a flow cell at an actin concentration of $1.4 \mu\text{g ml}^{-1}$ and allowed to bind to the surface for 30 s. BSA (0.5 mg ml^{-1}) in assay buffer was applied to wash unbound materials away and to block the free surface from non-specific protein binding. ATP (1 mM) in assay buffer was perfused through the flow cell to start the movement of the actin filaments. Fluorescence images were processed (averaging, linear contrast enhancement and superimposition) with a video-image processor (series 151, Imaging Technology) controlled by a microcomputer (IBM PC/AT).

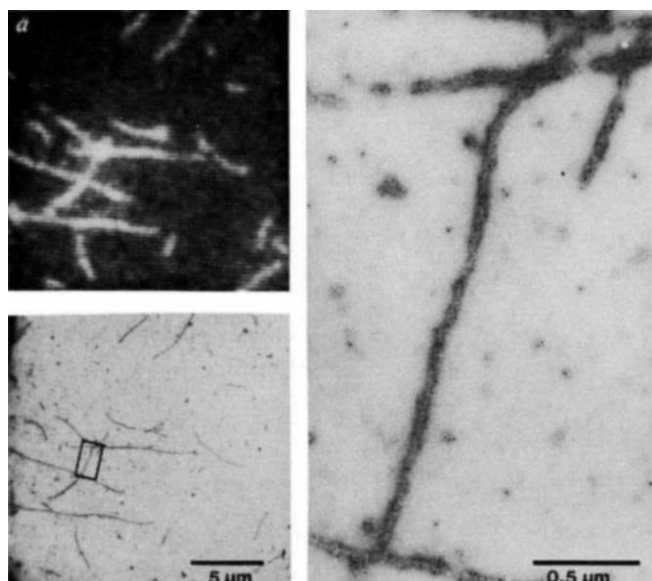
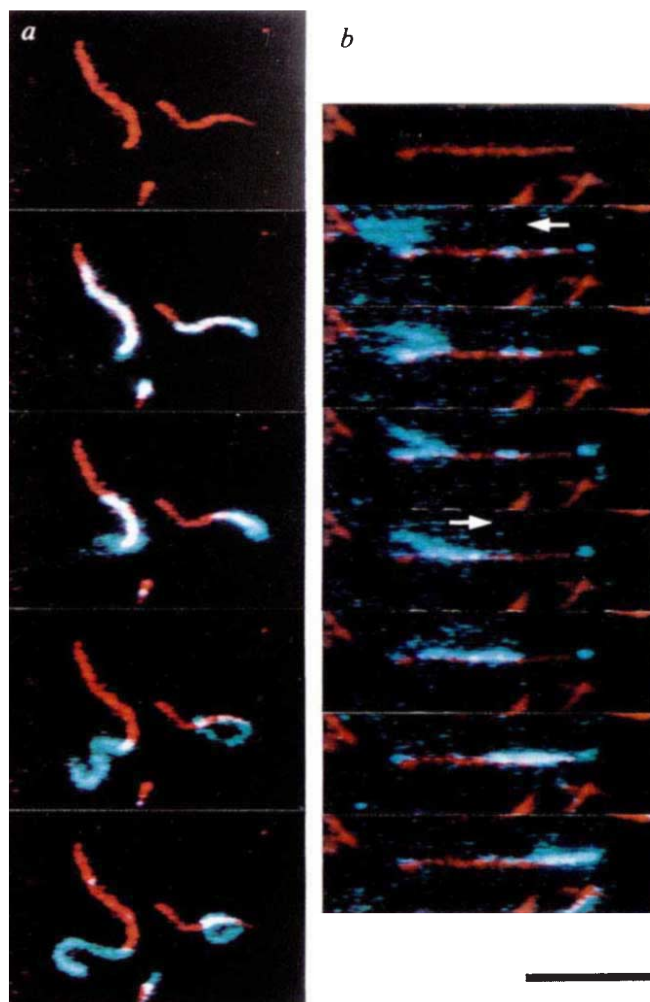
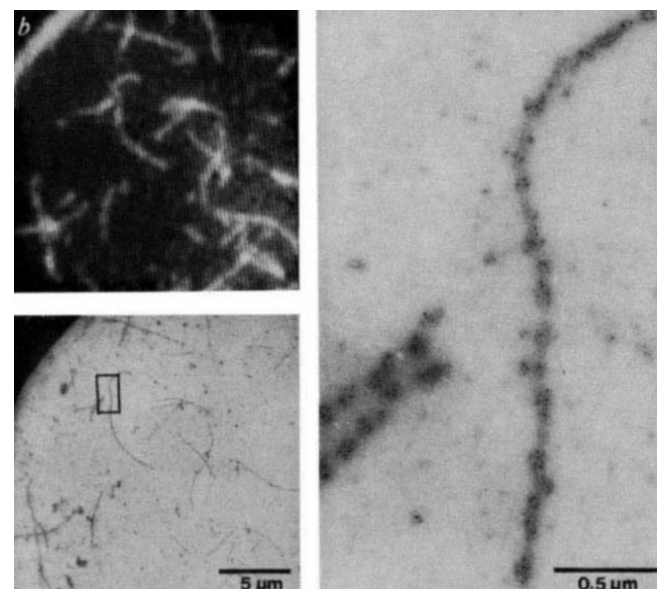


FIG. 2 Fluorescence images of actin filaments decorated with HMM and electron micrographs of the same fields. Fluorescence images (upper left) were recorded before ATP addition. Electron microscopic images of the same fields (lower left) were recorded before (*a*) or after (*b*) ATP addition; higher magnification images of boxed-off areas (right) show immunogold labelling of HMM. Note that the appearance of the HMM tracks (for example width and density of HMM) seemed to be unaltered by the movement of actin filaments. Scale bars, 5 μm (left) and 0.5 μm (right).

METHODS. To observe an identical field by both fluorescence and electron microscopy, an electron microscope finder grid (HF 7, Graticule, UK) was sandwiched between a coverslip and the nitrocellulose film that constituted the top part of the flow cell. Fluorescence images of actin filaments on marked grid squares were recorded first. Either before or after the addition of ATP, antibody against myosin subfragment-1 (25K-50K junction) was applied to the flow cell and incubated for 40 min. Unbound antibody was washed away with 0.5 mg ml^{-1} bovine serum albumin, and secondary anti-



body conjugated with gold (AuroProbe EM, Janssen, Belgium) was applied. After 40 min of the incubation and washing, the grid was picked up from the coverslip and negatively stained with 1% uranyl acetate. Electron microscope pictures of the marked grid squares were taken in an electron microscope 400T (Philips) at $\times 10,000$ and $\times 33,000$ magnification.

cence image of decorated actin filaments before ATP addition, and electron micrographs of the corresponding field. Filaments were well preserved and remained at the same locations despite the incubation with antibodies and the negative staining. More than 99% of the filaments were identified in both images (left). At a higher magnification (right), gold patches were seen to align along the actin filaments. The density of gold particles in the background reflects a random distribution of HMM which was not of sufficient density to support actin filament movement. Figure 2b shows a fluorescence image before ATP addition, and electron micrographs of the corresponding area after actin filaments had moved off the tracks in the presence of ATP. Gold particles still remained along the initial tracks of decorated filaments (right). The width of the track or the density of gold particles in the background did not increase appreciably. Furthermore, the density of the gold particles along the tracks remained almost the same (31.1 ± 2.8 particles per μm , $n = 20$ filaments, before ATP addition; 30.6 ± 3.4 particles per μm , $n = 23$, after ATP addition). These observations indicate that the HMM molecules associated with the nitrocellulose surface while maintaining unipolar binding to actin filaments, and stayed attached to the surface, probably in their original locations, even after actin filaments moved away.

To examine the flexibility of the nitrocellulose-bound HMM in binding to actin filaments, tracks of HMM were formed using long actin filaments, which were subsequently removed by ATP addition, and then very short actin filaments prepared with severin⁷ were introduced in the absence of ATP. If HMM could bind to actin filaments lying in either direction along the track, then it was anticipated that we should be able to see arrowhead patterns pointing in opposite directions on a single track of HMM. Figure 3b shows several short repeats of arrowheads

formed on a long track; they indeed show opposite polarities (see long arrows). Immunogold labelling of severin (Fig. 3, short arrows), which binds to the barbed end of actin filaments⁸, corroborates the polarity.

These experiments show that (1) actin filaments can follow the path pre-determined by HMM locations irrespective of the filament length, (2) actin filaments can move along the same HMM track in opposite directions, and (3) HMM bound to a nitrocellulose surface is flexible and can form a rigor complex even with an actin filament having opposite polarity to that of the initial filament. Two earlier observations showed that there is considerable flexibility in the HMM portion of the myosin molecule. First, two heads of HMM bind to actin molecules on the same strand in the absence of ATP⁹, showing that the two heads are translationally related in this bound configuration. Second, a single IgG molecule sometimes cross-links two heads of myosin^{10,11}, showing that the two heads can be arranged so that the same epitopes on the two heads face each other. Thus, there must be a considerable rotational freedom in the head-rod junction or head itself. The head-rod junction is known to have a large degree of freedom (for example, ref. 12); the myosin head may be rotating rapidly around its long axis looking for the correct binding site on the actin. One possible interpretation for the bidirectional movement described here is, therefore, that 180° rotational freedom exists in HMM. This interpretation is strongly supported by the complementary experiments reported by Reedy *et al.*¹³. They showed that myosin heads tethered in a single thick-filament of a mutated *Drosophila* muscle sarcomere can bind with opposite rigor-cross-bridge angles to flanking thin filaments, which are apparently of opposite polarities. □

Received 10 July; accepted 9 August 1989.

1. Yano, M., Yamamoto, Y. & Shimizu, H. *Nature* **299**, 557–559 (1982).
2. Hynes, T. R., Block, S. M., White, B. T. & Spudis, J. A. *Cell* **48**, 953–963 (1987).
3. Toyoshima, Y. Y. *et al.* *Nature* **328**, 536–539 (1987).
4. Kishino, A. & Yanagida, T. *Nature* **334**, 74–76 (1988).
5. Huxley, H. E. *J. molec. Biol.* **7**, 281–308 (1963).
6. Moore, P. B., Huxley, H. E. & DeRosier, D. J. *J. molec. Biol.* **50**, 279–295 (1970).
7. Yamamoto, K., Pardee, J. D., Reidler, J., Stryer, L. & Spudis, J. A. *J. Cell Biol.* **95**, 711–719 (1982).
8. Giffard, R. G., Weeds, A. G. & Spudis, J. A. *J. Cell Biol.* **98**, 1796–1803 (1984).
9. Craig, R. *et al.* *J. molec. Biol.* **140**, 35–55 (1980).
10. Winkelmann, D. A. & Lowey, S. *J. molec. Biol.* **188**, 595–612 (1986).
11. Miyazaki, T., Toyoshima, C., Wakabayashi, T. & Matsuda, G. *et al.* *J. Biochem., Tokyo* **103**, 458–462 (1988).
12. Kinoshita, K., Ishiwata, S., Yoshimura, H., Asai, H. & Ikegami, A. *Biochemistry* **23**, 5963–5975 (1984).
13. Reedy, M. C., Beall, C. & Fyberg, E. *Nature* **339**, 481–483 (1989).
14. Kron, S. J. & Spudis, J. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6272–6276 (1986).

ACKNOWLEDGEMENTS. We thank Dr K. Sutoh, University of Tokyo, for the gift of SI antibodies. This work was supported by a grant from NIH to J.A.S.; Y.Y.T. and C.T. were Postdoctoral Fellows of the Muscular Dystrophy Association.

Segment-specific expression of a homoeobox-containing gene in the mouse hindbrain

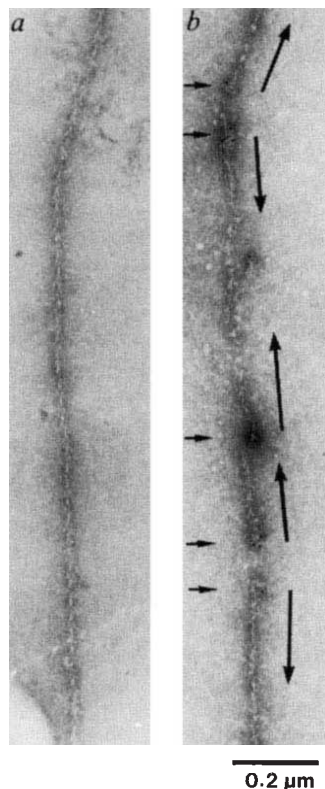
Paula Murphy, Duncan R. Davidson & Robert E. Hill

MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

THE process of segmentation, in which the developing embryo is divided into repetitive structures along its antero-posterior (A-P) axis, as a means of organizing and coordinating the body plan is found in a wide range of organisms. In *Drosophila*, homoeotic genes are involved in all levels of segmental organization and in determining segment identity¹. The roles of these genes in segmentation have been found mainly by mutational studies^{2,3}, but also by *in situ* hybridization, which has shown their domains of expression. In contrast to *Drosophila*, however, embryonic

FIG. 3 a, Negatively stained actin filament decorated with HMM on a nitrocellulose film of a flow cell; well-ordered arrowhead patterns show unipolar binding of HMM along the actin filament. b, Short decorated actin filaments pointing in opposite directions (long arrows) on a single long track of HMM, showing that the HMM bound to the nitrocellulose film could bind to actin filaments with either polarity. HMM tracks were first formed by placing onto the nitrocellulose long decorated actin filaments, which were removed later by addition of ATP; short actin filaments prepared with severin were then introduced in the absence of ATP to form rigor complexes with the HMM on the track. The gold particles (short arrows) show the presence of severin at the barbed ends of the short actin filaments, corroborating the polarity indicated by the arrowhead pattern. Scale bar, 0.2 μm .

METHODS. Short actin filaments were prepared with severin, an actin-filament severing protein⁷. Severin was first mixed with actin labelled with rhodamine-phalloidin at a 1:100 molar ratio in the absence of Ca^{2+} ; 0.1 mM Ca^{2+} was subsequently added to the mixture to activate the severin⁷. Antibody against severin and gold-conjugated secondary antibody were applied after introduction of the actin filaments into a flow cell, and the specimens were negatively stained with 1% uranyl acetate.



expression of homoeobox-containing genes in vertebrate organisms has not been found to follow a segmental pattern⁴. Vertebrate segmentation can be clearly seen in the mesodermal somites, but repetitive morphological structures in the central nervous system (neuromeres)⁵⁻¹³ have only recently been shown to have developmental significance. Neuromeres in the hindbrain (rhombomeres) have been defined as segmental units by their pattern of nerve formation in the developing chick¹⁴ and by the alternating expression of *Krox-20*, a gene encoding a zinc-finger DNA-binding protein, in the 9.5-day-old mouse¹⁵. Here we report that a mouse homoeobox-containing gene, *Hox-2.9*, is expressed in a segment-specific manner in the developing mouse hindbrain. This expression is in a region which is flanked by the regions of expression of *Krox-20*, and is precisely contained within a single neuromere, rhombomere 4.

A homoeobox-containing gene mapping to the *Hox-2* gene cluster on chromosome 11 (refs 16 and 17), which we have characterized as *Hox-2.9* (P.M. *et al.*, manuscript in preparation; R. Krumlauf, personal communication), was isolated from an 8.5-day-old mouse embryonic complementary DNA library. We determined the pattern of *Hox-2.9* expression during develop-

ment by *in situ* hybridization using a probe specific for *Hox-2.9*, 3' to the homoeobox, and analysing serial sections from 6.5-day-, 8.5-day-, 9.5-day-, 10-day- and 13.5-day-old embryos, using at least four embryos from each stage.

Hox-2.9 transcripts were not detected in the 6.5-day-old embryo, when gastrulation begins. At 8.5 days of development the embryo is at the neural fold stage. At this stage, expression of *Hox-2.9* is detected in both the neural ectoderm and the underlying mesoderm (Fig. 1*a, b*). The domain of expression is extensive, with an anterior limit in the developing hindbrain posterior to the preotic sulcus (Fig. 1*c, d*). The low level of expression and the absence of morphological markers in the hindbrain, however, made it difficult to locate the anterior limit of expression more precisely. Mesodermal expression was restricted to lateral and presomitic mesoderm; no label was detectable in the somites.

By 9.5 days the expression pattern had changed dramatically, as had the complexity of the embryo. At this stage the segmental units of the hindbrain, the rhombomeres, could easily be seen as a series of neuroepithelial swellings in frontal and parasagittal sections (Figs 2 and 3*d*). Individual rhombomeres could be

FIG. 1 *In situ* hybridization of a ³⁵S-labelled anti-sense probe for *Hox-2.9* to 8.5-day-old mouse-embryo sections. *a, b*, Sagittal section showing expression in the mesoderm (m) and neural ectoderm (ne). *c, d*, Sagittal section through the headfold (hf) region showing the anterior limit of expression in the neuroectoderm and lateral mesoderm (arrow). *a* and *c*, Brightfield photographs; *b* and *d* are the corresponding darkfield photographs. s, Somite; pos, preotic sulcus.

METHODS. The *in situ* hybridization procedure used was as previously described³¹, with $\sim 4 \times 10^5$ d.p.m. of ³⁵S-labelled UTP incorporated into RNA applied per section. Sense-strand control probe was included routinely and showed no hybridization signal above background (data not shown).

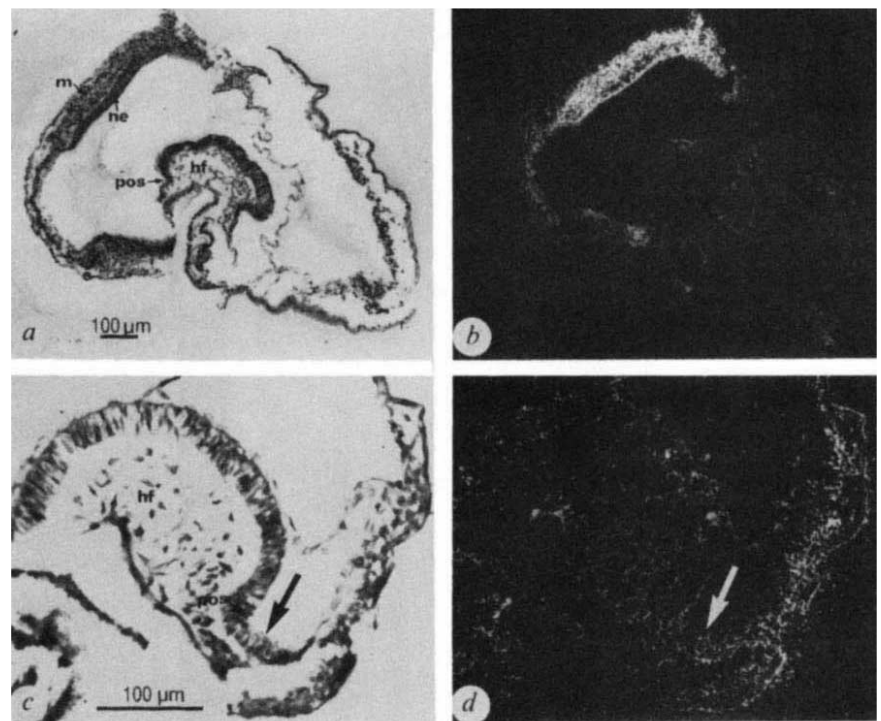
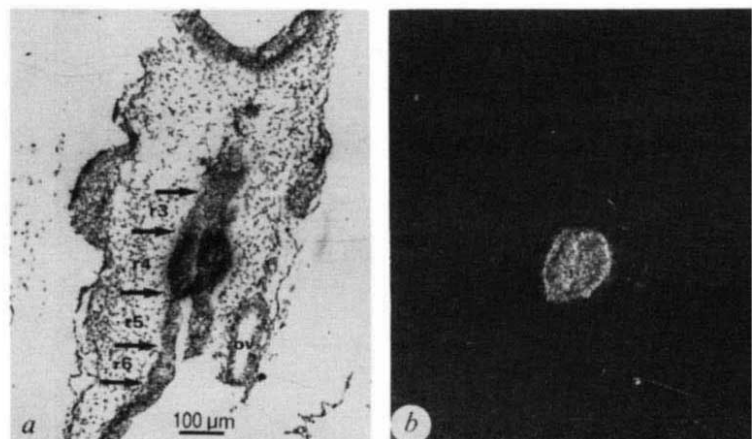


FIG. 2 Frontal section through the hindbrain of a 9.5-day-old embryo showing localized expression of *Hox-2.9* within rhombomere 4 (r4). *a*, Brightfield photograph. *b*, Corresponding darkfield photograph. ov, Otic vesicle; arrows indicate neuromere boundaries; rhombomeres are numbered r3, r4, r5, r6.



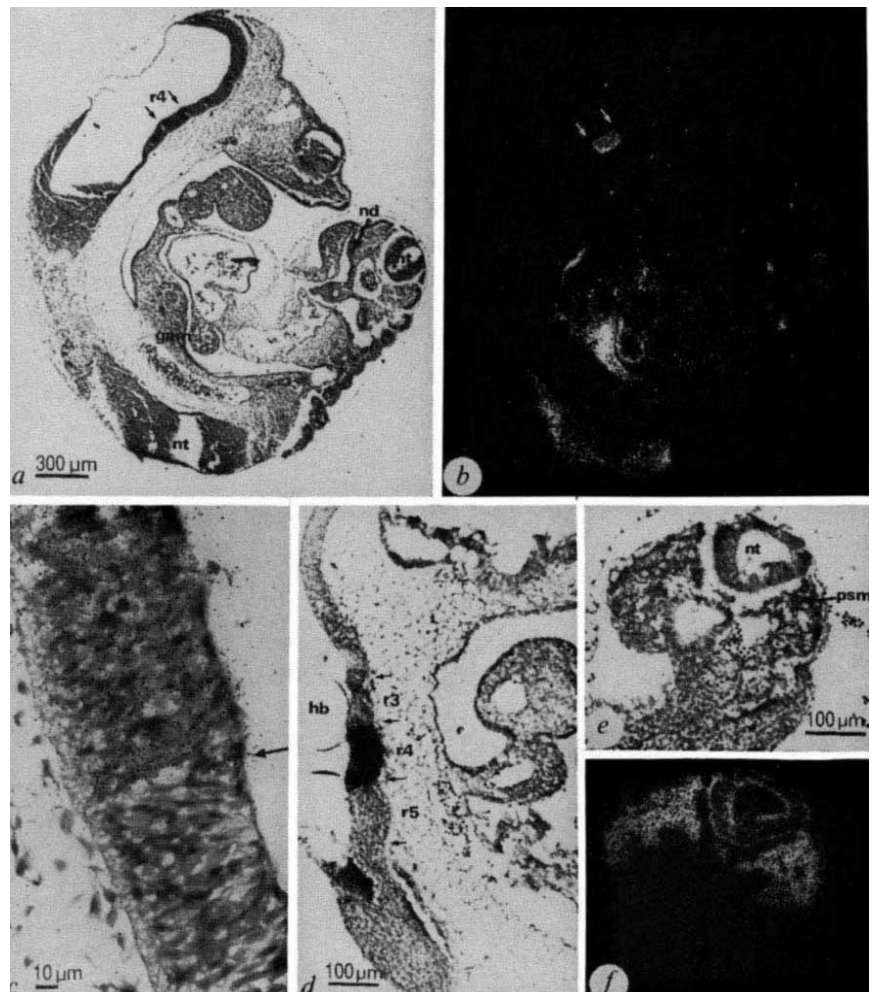
identified visually by their position relative to the otic vesicle and numbered as shown in Fig. 4. In the hindbrain, *Hox-2.9* was expressed at this stage in a single segment, rhombomere 4. It was difficult, however, to relate rhombomere 4 to the hindbrain expression at day 8.5, when rhombomeres are not morphologically detectable. Expression was detected throughout rhombomere 4, with the exception of the thin roof plate. The boundaries of expression corresponded precisely with the morphological boundaries of rhombomere 4 (Figs 2 and 3d), and Fig. 3c shows the precision of this boundary at the cellular level. Rhombomere 4 seems to be the principal site of expression of *Hox-2.9*, because the very localized signal seen at this site was consistently much higher than any other signal detected.

Outside the hindbrain the anterior neural tube was unlabelled at 9.5 days, but expression was detected in the neural tube posterior to the forelimb buds (Figs 3a, b). Expression was still detected in the presomitic mesoderm but not in the somites (Figs 3e, f). Gut-associated mesenchyme and the epithelium of the upper gut posterior to the third branchial arch were also labelled (Fig. 3a, b), the gut-associated-mesenchyme having been derived from lateral-plate mesoderm, which showed expression at day 8.5. A low level of expression was detected in the nephrogenic duct (Fig. 3a, b). Analysis of sections through 10-day-old embryos showed no further change in *Hox-2.9* expression. Unlike most homeobox-containing genes previously analysed, the expression of *Hox-2.9* did not persist in the late embryo stage: at 13.5 days of development, expression could not be detected in any part of the embryo by examination of serial sections in sagittal, transverse or frontal planes. Neuromeric boundaries could also no longer be seen at this stage.

The expression of *Hox-2.9* in rhombomere 4 indicates that this gene plays a part in the formation or differentiation of this segment. Neuromeres have segmental identities, as seen by the exit points of cranial ganglia from specific neuromeres¹⁴ with the facial and acoustic ganglia extending from rhombomere 4. Independent evidence for the genetic control of development within this neuromere comes from the phenotype of the mouse developmental mutant *kreisler*, which displays faulty segmentation of the hindbrain and specific degeneration of the cells of rhombomere 4 by day 9.5 (ref. 18). The mutation maps to mouse chromosome 2 (ref. 19) and is therefore not allelic with *Hox-2.9*. It is possible that *Hox-2.9* and *kreisler* are both genetic regulators of segmentation in the hindbrain. We can now investigate the relationship between these two genes.

Krox-20 is expressed in rhombomeres 3 and 5 (ref. 15), which flank the region of expression of *Hox-2.9* at day 9.5. At day 8.5, before the neuromeres become visible, *Krox-20* expression is first detected in the region of two neuroepithelial invaginations, proneuromeres-A and -B, indicating that segments start to become established at this stage¹⁵. *Krox-20* is therefore a likely candidate for involvement in the process of segmentation because its expression precedes the appearance of morphological segments. The spatial and temporal correlation between the expression of *Hox-2.9* and *Krox-20* raises the possibility that their products interact, directly or indirectly, in the process of segmentation in this region of the central nervous system (CNS). Both of these genes are thought to encode regulatory proteins on the basis of the DNA-binding motifs that the proteins possess: *Hox-2.9* encodes a protein that has a helix-turn-helix motif²⁰ and *Krox-20* encodes a protein with a zinc-finger DNA-binding

FIG. 3 Expression of *Hox-2.9* in 9.5–10-day-old mouse embryos. a, b, Sagittal section of a 10-day-old embryo showing expression in the hindbrain (r4, rhombomere 4), the gut-associated mesenchyme (gam), the nephrogenic duct (nd), and the posterior neural tube (nt). c, Boundary between r3 and r4 (marked with arrow) in a 10-day sagittal section showing expression of *Hox-2.9* within r4. The short autoradiographic exposure time (two weeks) allowed the labelling of individual cells to be observed and emphasized the sharpness of this expression boundary. d, Sagittal 9.5-day embryo section probed with the *Hox-2.9*-specific probe and exposed for a long period (5 weeks) shows the morphology of the hindbrain and the expression of *Hox-2.9* in r4. hb, Hindbrain. e, f, Transverse section showing expression of *Hox-2.9* in presomitic mesoderm (psm) and the neural tube (nt) in posterior regions of the 9.5-day-old embryo. a and e, Bright field photographs. b, f, Corresponding darkfield photographs.



motif^{21,22}. There is evidence of interaction between proteins with these regulatory domains within the *Drosophila* segmentation system²³.

An evolutionary relationship has been suggested between genes of the *Hox-2* and *Hox-1* clusters in the mouse and the homoeotic genes of the Antennapedia complex in *Drosophila*^{16,17}. The sequence of the *Hox-2.9* homoeobox (data not shown) shows a high degree of similarity to *Hox-1.6*, which is most similar to the *Drosophila* gene *labial*²⁴. It is of interest that both *Hox-2.9* and *Hox-1.6* have the same temporal pattern of expression²⁵ (P.M. *et al.*, manuscript in preparation). *Labial* has a homoeotic role in determining segmental identity and its pattern of expression marks a single ancestral segment in the developing *Drosophila* head^{26,27}. Most evidence suggests that segmentation arose independently in invertebrates and vertebrates²⁸; the expression of *labial* and *Hox-2.9* in single segments may reflect parallels in the mechanisms that determine the identity of anterior segments in invertebrates and vertebrates.

The results presented here raise the possibility that other homoeobox-containing genes are involved in specifying neural segments. Numerous expression studies on mouse homoeobox-containing genes show that their transcripts are most abundant within the developing CNS, where they occupy large overlapping domains⁴. Several of these transcripts have anterior boundaries to their expression in the hindbrain¹⁶, and our work indicates that studies should now be undertaken to identify these boundaries in sections where neuromeres are well defined. The distinct anterior limit of *Hox-1.5* expression lies at a neuromere boundary²⁹. We infer, from its position rostral to the otic vesicle³⁰, that this boundary lies between rhombomeres 4 and 5, which coincides with the posterior limit of *Hox-2.9* expression.

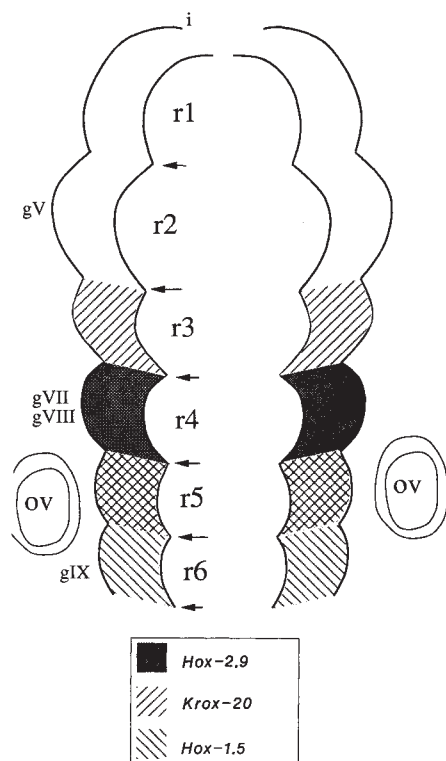


FIG. 4 Diagrammatic representation of a frontal section through the hindbrain of a 9.5-day mouse embryo showing the position of the neuromeres (rhombomeres) with respect to the otic vesicle (ov), the hindbrain-midbrain junction (isthmus, i), and the cranial ganglia which are represented alongside the rhombomeres from which they extend (gV–gIX). The arrows indicate rhombomere boundaries, and the rhombomeres are numbered from anterior to posterior (r1–r6). The shaded areas indicate the domains of expression of *Hox-2.9*, *Hox-1.5* and *Krox-20* in the hindbrain (see text for refs). *Hox-1.5* expression extends posteriorly into the neural tube.

It has been suggested that the overlapping domains of homoeobox-containing-gene expression convey positional information in the form of unique combinations of gene products in blocks along the A–P axis⁴. It remains to be seen if these blocks relate to neuromeres. In this context, *Hox-2.9* is unique in being expressed within a single neuromere, and its expression within rhombomere 4 indicates that it is involved in specifying the unique identity of this segment. □

Received 24 May; accepted 4 August 1989.

1. Akam, M. *Development* **101**, 1–22 (1987).
2. Lewis, E. B. *Nature* **276**, 565–570 (1978).
3. Kaufman, T. C. in *Time, Space and Pattern in Embryonic Development*. 365–383 (Liss, New York, 1983).
4. Holland, D. W. D. & Hogan, B. L. M. *Genes Dev.* **2**, 773–782 (1988).
5. Orr, H. J. *Morph.* **1**, 311–372 (1887).
6. Streeter, G. L. *Anat. Rec.* **2**, 111–115 (1908).
7. Neal, H. V. J. *Morph.* **31**, 293–315 (1918).
8. Adelman, H. B. J. *comp. Neurol.* **39**, 19–171 (1925).
9. Berquist, H. & Kallen, B. J. *comp. Neurol.* **100**, 627–660 (1954).
10. Vaage, S. *Adv. Anat. Embryol. Cell Biol.* **41**, 1–88 (1969).
11. Tuckett, F., Lim, L. & Morriss-Kay, G. M. J. *Embryol. exp. Morph.* **87**, 215–228 (1985).
12. Tuckett, F. & Morriss-Kay, G. M. J. *Embryol. exp. Morph.* **88**, 231–249 (1985).
13. Sakai, Y. *Anat. Rec.* **218**, 450–457 (1987).
14. Lumsden, A. & Keynes, R. *Nature* **337**, 424–428 (1988).
15. Wilkinson, D. G., Bhatt, S., Chavir, P., Bravo, R. & Charnay, P. *Nature* **337**, 461–464 (1989).
16. Graham, A., Papalopulu, N. & Krumlauf, R. *Cell* **57**, 367–378 (1989).
17. Do, M.-S. & Lomax, P. *Genomics* **3**, 195–200 (1988).
18. Deol, M. S. J. *Embryol. exp. Morph.* **12**, 475–490 (1964).
19. Hertwig, P. Z. *Indukt. Abstammungs-Vererbungsl.* **80**, 220–246 (1942).
20. Laughon, A. & Scott, M. P. *Nature* **310**, 25–31 (1984).
21. Miller, J., McLachlan, A. D. & Klug, A. *EMBO J.* **4**, 1609–1614 (1985).
22. Chavir, P. *et al.* *EMBO J.* **7**, 29–35 (1988).
23. Ingham, P. *Nature* **335**, 25–34 (1988).
24. Miodzik, M., Fjose, A. & Gehring, W. J. *EMBO J.* **7**, 2569–2578 (1988).
25. Duboule, D. & Dolle, P. *EMBO J.* **8**, 1497–1505 (1989).
26. Diederich, R. J., Merrill, V. K. L., Pultz, M. A. & Kaufmann, T. C. *Genes Dev.* **3**, 399–414 (1989).
27. Mahaffey, J. W., Diederich, R. J. & Kaufmann, T. C. *Development* **105**, 167–174 (1989).
28. Hogan, B., Holland, P. & Schöfield, P. *Trends Genet.* **1**, 67–74 (1985).
29. Gaunt, S. J. *Development* **101**, 51–60 (1987).
30. Gaunt, S. J., Miller, J. R., Powell, D. J. & Duboule, D. *Nature* **324**, 662–664 (1986).
31. Davidson, D., Graham, E., Sime, C. & Hill, R. *Development* **104**, 305–316 (1988).

ACKNOWLEDGEMENTS. We thank Elizabeth Graham for help with *in situ* hybridizations, Norman Davidson for photographic expertise, Christine Sime for help in cloning *Hox-2.9*, David Porteous and Nick Hastie for critical reading of the manuscript, and Gillian M. Morriss-Kay for interpretation of early embryo sections.

CD3-negative natural killer cells express ζ TCR as part of a novel molecular complex

Paul Anderson^{*†}, Michael Caligiuri^{*}, Jerome Ritz^{*} & Stuart F. Schlossman^{*}

^{*} Division of Tumor Immunology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA

[†] Department of Rheumatology and Immunology, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, Massachusetts 02115, USA

NATURAL killer (NK) cells are large granular lymphocytes capable of killing tumour cells in a non-MHC restricted manner^{1,2}. NK cells do not express cell-surface CD3, or any known target recognition structure analogous to the T cell antigen receptor (TCR) heterodimers ($\alpha\beta$ or $\gamma\delta$)^{3–8}. Consistent with their lack of expression of a CD3–TCR complex, NK cells do not require prior sensitization or antigen presentation by accessory cells to specifically recognize their tumour targets¹. Although NK cells do not express CD3–TCR, they do express CD2, the target of an alternative activation pathway which is functional in both T cells and NK cells^{9–12}. In T cells, this alternative activation pathway utilizes some component of the CD3–TCR complex as a transducer molecule that is required for mitogenesis^{13–15}. The fact that NK cells are activated by this alternative pathway suggested that they might express a related subunit of the CD3–TCR complex capable

of transducing the CD2-mediated signal. Here we show that human NK cells express the ζ -chain of the TCR complex in association with additional structures not included in CD3-TCR.

Total cellular RNA extracted from clonal and polyclonal populations of human peripheral blood NK cells was probed for its expression of ζ TCR messenger RNA by northern blotting (Fig. 1, top). Compared with CD3⁺ T cells (lane 1), both the clonal (JJ1, lane 2) and polyclonal (lane 3) NK populations expressed significant amounts of ζ -chain message. By contrast, neither the JJ1 NK clone, nor the polyclonal population of peripheral blood CD3⁺ NK cells expressed detectable amounts of CD3 γ mRNA (Fig. 1, middle). As previously reported⁷, both NK populations expressed small amounts of CD3 ϵ mRNA (Fig. 1, bottom panel). These results suggested that one or both of these TCR subunits might participate in a novel NK cell receptor complex.

To determine whether either of these mRNAs was translated into a stably expressed protein product, we screened two CD3⁺

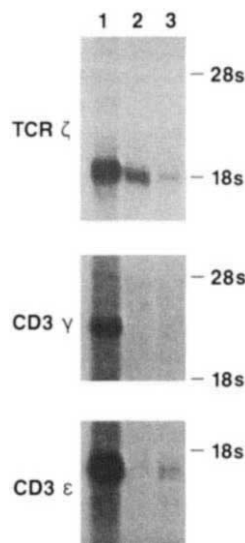


FIG. 1 Northern blot analysis of T cells and NK cells. Lanes 1, 2, and 3 contain RNA extracted from a CD4⁺ T-cell clone (T4Cl), an NK clone (JJ1) and peripheral blood CD3⁺ NK cells, respectively. Nitrocellulose blots were probed with complementary DNA inserts encoding ζ TCR (top), CD3 γ (middle) and CD3 ϵ (bottom) as described below.

METHODS. Cellular RNA was extracted²³ from a CD4⁺ T-cell clone (T4Cl), an NK clone (JJ1), and peripheral-blood NK cells prepared in the following manner. Peripheral blood lymphocytes were isolated from whole blood by centrifugation over Ficoll Hypaque. T lymphocytes were depleted by antibody-mediated complement lysis using a monoclonal antibody reactive with CD6. The remaining cells were cultured in the presence of recombinant IL-2 (1,000 U ml⁻¹) for 14 days. Viable cells were collected by centrifugation over Ficoll Hypaque and further depleted of cells expressing CD3 using magnetic beads coupled to anti-CD3 antibody. The resulting population was phenotypically <1% CD3⁺ and 89% NKH1⁺, and demonstrated non-MHC restricted killing of tumour targets as previously described²⁴. RNA from CD3⁺ T cells (5 μ g poly(A)⁺ RNA, lane 1), the JJ1 NK clone (20 μ g total cellular RNA, lane 2), and polyclonal CD3⁺ NK cells (20 μ g total cellular RNA, lane 3), was separated on a 1.2% formaldehyde agarose gel, transferred to nitrocellulose and probed for the expression of the indicated mRNAs. A full length cDNA encoding the human TCR ζ -chain in the vector pGEM3 was a gift from A. Weissman²⁵. Insert DNA was excised using appropriate restriction endonucleases, separated on a 1% low melt agarose gel, and extracted with phenol. Insert DNA was labelled with ³²P by nick translation and hybridized to the nitrocellulose blot described above. The nitrocellulose blot hybridized to ζ TCR was stripped and rehybridized using cDNA probes encoding CD3 γ (ref. 26) and CD3 ϵ (ref. 27), both gifts from C. Terhorst. The CD3 γ mRNA is detected as a 3.5 kilobase (kb) transcript and the CD3 ϵ mRNA is detected as a 1.8 kb transcript. The relative migration of 28S and 18S ribosomal RNAs are indicated.

peripheral blood NK cell clones for their expression of CD3 ϵ CD2, ζ TCR and CD56 (NKH1) by flow cytometry. Figure 2 shows the flow cytometry phenotype of two CD3⁺ NK clones that have been described previously¹. Both clones were found to be CD3⁺, CD2⁺, ζ TCR⁺ and NKH1⁺. The presence of ζ TCR in clonal and polyclonal populations of CD3⁺ NK cells was confirmed by the immunoblotting analysis shown in Fig. 3. Detergent lysates prepared from T cells (lane 1), NK clones JJ1 and JT_B18 (lanes 2 and 3) and interleukin-2 (IL-2) expanded peripheral blood CD3⁺ NK cells (lane 4) all expressed significant amounts of ζ TCR migrating as a single band with a

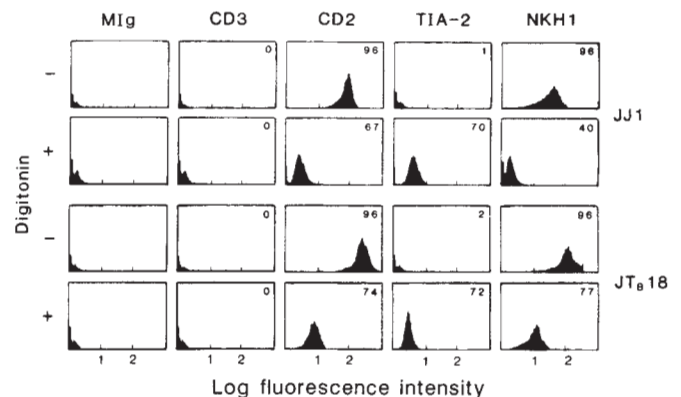


FIG. 2 Flow cytometric analysis of permeabilized and unpermeabilized NK clones.

METHODS. Clonal populations of peripheral blood NK cells were isolated as described²⁸. These cells were capable of non-MHC restricted cytotoxicity against a variety of tumour targets as has been reported^{1,28}. Clones designated JJ1 and JT_B18 were stained with control ascites (Mig), monoclonal antibodies reactive with CD2 (ref. 9) (1 old 24Cl, immunoglobulin (IgG2), CD3 ϵ (refs 29,30) (RW24B6, IgG2), ζ TCR (ref. 20) (TIA-2, IgG1) or NKH1 (ref. 31) (IgG1), and analysed by flow cytometry. Both clones were stained in the absence or presence of digitonin, a permeabilization agent that allows the detection of intracellular targets by exogenously added antibody³². The monoclonal antibody reactive with ζ TCR (TIA-2) has been shown to stain permeabilized, but not unpermeabilized T cells, suggesting that it recognizes an intracellular epitope of this antigen receptor subunit²⁰. Individual histograms are derived from analysis of 10,000 individual cells. Numbers in the upper right-hand corner of each histogram give the percentage of positively staining cells.

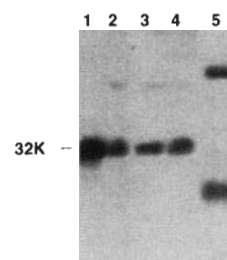


FIG. 3 Immunoblotting analysis using a monoclonal antibody reactive with ζ TCR (TIA-2). Lane 1, Peripheral blood T lymphocytes; Lane 2, NK clone JJ1; Lane 3, NK clone JT_B18; Lane 4, peripheral blood NK cells; Lane 5, 1 μ g of IgG1 antibody treated with 2-mercaptoethanol. Heavy and light chains are visualized by virtue of their reactivity with the rabbit anti-mouse developing reagent and serve as size markers.

METHODS. The indicated cell types (5 $\times$ 10⁶ cells per sample) were solubilized in NP-40 lysis buffer (1% NP-40, 150 mM NaCl, 1 mM EDTA, 1 mM PMSF, 50 mM Tris, pH 8.0) for 30 min on ice, then centrifuged for 30 min at 12,000g. Supernatants were diluted 1:1 with sample buffer (2% SDS, 10% glycerol, 0.1 M Tris HCl, pH 6.8, 0.02% bromophenol blue), separated on 10% SDS polyacrylamide gels, transferred to nitrocellulose, and probed with TIA-2 as previously described²⁰.

relative molecular mass (M_r) of 32,000 (32K). In some immunoblots, ζ TCR migrated as a doublet with the upper band at 32K. The relative expression of ζ TCR in CD3⁺ T cells and NK cells was estimated from both northern analysis (normalized to the relative expression of actin mRNA, not shown) and immunoblotting analysis (Fig. 3) to be ~2:1. These results indicate that ζ -chain expression is not limited to cells possessing the entire CD3-TCR complex.

In T cells, the ζ -chain appears to regulate signal transduction mediated by the antigen-receptor heterodimer¹⁶⁻¹⁹. If it has a similar function in NK cells, its presence might explain the differential responsiveness of T cells and NK cells to CD2 mediated activation in the absence of cell-surface CD3. By comparing the ζ TCR-containing complexes immunoprecipitated from digitonin lysates prepared from T cells to those immunoprecipitated from lysates prepared from NK cells, we hoped to identify candidate NK receptor complexes that might be functionally analogous to the T-cell-antigen-recognition unit. Digitonin-solubilized cells were immunoprecipitated with antibodies reactive with either CD3 ϵ or ζ TCR, and analysed on non-reducing/reducing diagonal gels as shown in Fig. 4. As we have previously reported²⁰, peripheral blood T lymphocytes express a receptor complex composed of an $\alpha\beta$ heterodimer, the CD3 molecular complex (γ , δ and ϵ), and ζ (Fig. 4a). The receptor complexes immunoprecipitated by anti-CD3 ϵ and ζ TCR antibodies differed in two ways. First, ζ TCR immunoprecipitates included a 60-70K spot above the diagonal that was less prominent in CD3 ϵ immunoprecipitates. Second, ζ TCR immunoprecipitates included a novel heterodimer composed of ζ and a 12K structure that was not seen in CD3 ϵ immunoprecipitates. Although the 12K structure is relatively faint in the exposure shown in Fig. 4a, it was more prominent in longer exposures of this same autoradiogram (data not

shown). In contrast to T lymphocytes, anti-CD3 ϵ immunoprecipitates prepared from NK cell lysates did not contain any component of the TCR complex (Fig. 4b, c). In spite of the fact that these cells express small amounts of CD3 ϵ mRNA, they do not contain detectable amounts of CD3 ϵ protein. Consistent with the flow cytometry and immunoblotting data shown in Figs 2 and 3, immunoprecipitates formed using the ζ TCR reactive antibody TIA-2 contained significant amounts of ζ migrating as two distinct 16K spots off the diagonal. The migration patterns of proteolytic digests prepared from these 16K proteins were identical (data not shown), suggesting that in NK cells, ζ is expressed, in approximately equal amounts, as a disulphide-linked homodimer and as a disulphide-linked heterodimer with the 12K structure mentioned above (arrow, Fig. 4b, d). We were unable to obtain sufficient quantities of the 12K protein to compare it to ζ by proteolytic mapping. It is possible that the 12K protein arises from a proteolytic cleavage of one of the chains of the ζ homodimer. Additional experiments will be required to determine whether these disulphide-linked subunits are structurally related.

In addition to ζ , higher molecular weight structures were also included in these immunoprecipitates. In NK clone JJ1 (Fig. 4b), this structure migrated as a 60-70K spot above the diagonal that appears to be the same structure identified in T-cell lysates (Fig. 4a). In NK clone JT_B18, this structure migrated as an 80-90K on diagonal spot (Fig. 4c). Interestingly, polyclonal populations of CD3⁺ peripheral blood NK cells expressed a ζ complex that appeared to associate with both of the higher molecular weight forms identified in these clones (Fig. 4d). Because of their non-covalent association with ζ , we cannot be certain that these structures retain their association with ζ in the cell. If they are functionally important components of the ζ receptor complex on NK cells, the fact that they can also be

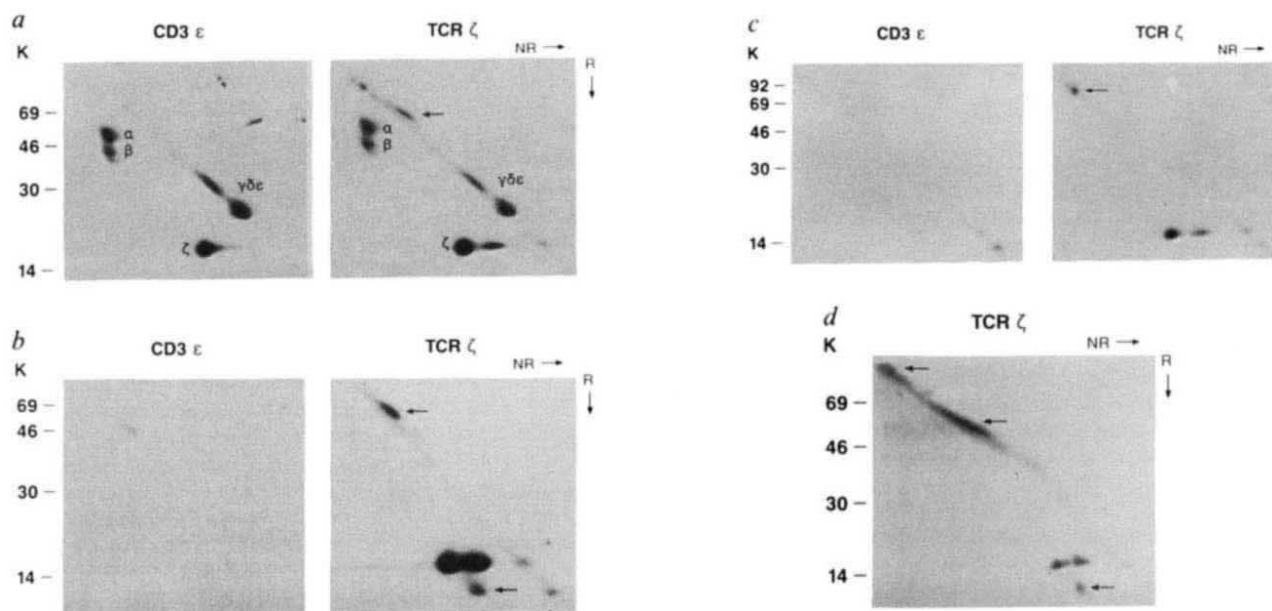


FIG. 4 Immunoprecipitation analysis using non-reducing/reducing diagonal gel electrophoresis. Left panels, CD3 ϵ immunoprecipitates. Right panels, ζ TCR immunoprecipitates. a, Immunoprecipitates from peripheral blood T-cell lysates; b, Immunoprecipitates from NK-clone JJ1; c, Immunoprecipitates from NK-clone JT_B18; d, Immunoprecipitates from peripheral-blood NK cells prepared as described in the legend to Fig. 1. The relative migration of molecular size markers are indicated on the left. Arrows point to the 12K ζ associated structure (b and d), the 60-70K ζ associated structure (a, b and d) and the 80-90K ζ associated structure (c and d).

METHODS. Peripheral blood T lymphocytes, NK clones JJ1 and JT_B18, and peripheral blood NK cells were permeabilized with digitonin, radiolabelled, and immunoprecipitated from digitonin lysis buffer (1% digitonin, 0.12%

Triton X-100, 150 mM NaCl, 1 mM PMSF, 20 mM triethanolamine, pH 7.8) using monoclonal antibodies reactive with CD3 ϵ (Leu 4) (ref. 33) and ζ TCR (TIA-1) as previously described²⁰. Because ζ TCR is poorly labelled by surface iodination, cells were routinely permeabilized before being radiolabelled. When unpermeabilized lymphocytes were radioiodinated and immunoprecipitated as described above, some labelling of ζ TCR could be obtained, and the ζ containing complex was similar to that observed using permeabilized cells. Electrophoretic analysis by non-reducing/reducing diagonal gel electrophoresis was as described²⁰. The first dimension, run under non-reducing (NR) conditions is depicted in the horizontal direction, whereas the second dimension, run under reducing (R) conditions is depicted in the vertical direction.

found in T cells is of great interest. It has been known for some time that a subpopulation of CD3⁺ T lymphocytes is capable of NK-like, non-MHC restricted killing of tumour targets^{21,22}. It would, therefore, not be surprising if a subpopulation of both NK cells and T cells shared a receptor structure capable of mediating this effect.

The expression of ζ TCR in CD3⁻ NK cells has important implications for the activation and function of these cells. Recent evidence suggests that ζ TCR plays a pivotal role in T-cell activation initiated through the T-cell receptor complex. If ζ plays a similar role in NK cells, the fact that it is found in association with additional structures not part of the CD3-TCR complex may be very important. Our results suggest that in both T cells and NK cells, ζ can exist as a disulphide-linked homodimer of 32K or as a disulphide-linked heterodimer with a 12K structure. In NK cells, ζ is also associated with additional higher molecular weight structures that appear to differ on individual clones. Although additional experiments will be necessary to fully elucidate the role of the NK cell ζ complex in non-MHC restricted killing, the identification of a ζ containing molecular complex in NK cells immediately suggests an explanation for some of the functional properties of these cells, and points to the requirement for future research aimed at understanding the receptor complex used by this lymphocyte subset. □

Received 30 May; accepted 7 August 1989.

1. Ritz, J., Schmidt, R. E., Michon, J., Hercend, T. & Schlossman, S. F. *Adv. Immun.* **42**, 181–211 (1988).
2. Hercend, T. & Schmidt, R. E. *Immun. Today* **9**, 291–293 (1989).

3. Lanier, L. L., Le, A. M., Phillips, J. H., Warner, N. L. & Babcock, G. F. *J. Immun.* **131**, 1789–1796 (1983).
4. Ritz, J. *et al. Science* **228**, 1540–1543 (1985).
5. Lanier, L. L., Cwirla, S. & Phillips, J. H. *J. Immun.* **137**, 3375–3377 (1986).
6. Lanier, L. L., Cwirla, S., Federspiel, N. & Phillips, J. H. *J. exp. Med.* **163**, 209–214 (1986).
7. Blassoni, R., Ferrini, S., Prigione, I., Moretta, A. & Long, E. O. *J. Immun.* **140**, 1685–1689 (1988).
8. Biron, C. A. *et al. J. Immun.* **139**, 1704–1708 (1987).
9. Meuer, S. C. *et al. Cell* **36**, 897–906 (1984).
10. Schmidt, R. E. *et al. J. Immun.* **135**, 672–677 (1985).
11. Siliciano, R. F., Pratt, J. C., Schmidt, R. E., Ritz, J. & Reinherz, E. L. *Nature* **317**, 428–430 (1985).
12. Schmidt, R. E. *et al. Nature* **318**, 289–291 (1985).
13. Breitmeyer, J. B., Daley, J. F., Levine, H. B. & Schlossman, S. F. *J. Immun.* **139**, 2899–2905 (1987).
14. Alcover, A., Chang, H. C., Sayre, P. H., Hussey, R. E. & Reinherz, E. L. *Eur. J. Immun.* **18**, 363–367 (1988).
15. Alcover, A. *et al. EMBO J.* **7**, 1973–1977 (1988).
16. Weissman, A. M., Samelson, L. E. & Klausner, R. D. *Nature* **324**, 480–482 (1986).
17. Samelson, L. E., Patel, M. D., Weissman, A. M., Harford, J. B. & Klausner, R. D. *Cell* **46**, 1083–1090 (1986).
18. Mercep, M. *et al. Science* **242**, 571–574 (1988).
19. Baniyash, M., Garcia-Morales, P., Luong, E., Samelson, L. E. & Klausner, R. D. *J. biol. Chem.* **263**, 18225–18230 (1988).
20. Anderson, P., Blue, M. L., O'Brien, C. & Schlossman, S. F. *J. Immun.* (in the press).
21. Hercend, T. *et al. J. exp. Med.* **158**, 1547–1560 (1983).
22. Schmidt, R. E., Murray, C., Daley, J. F., Schlossman, S. F. & Ritz, J. *J. exp. Med.* **164**, 351–356 (1986).
23. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning: A Laboratory Manual* 187–194 (Cold Spring Harbor Laboratory, 1982).
24. Schmidt, R. E. *et al. J. clin. Invest.* **79**, 305–308 (1987).
25. Weissman, A. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 9709–9713 (1988).
26. Krissansen, G. W., Owen, M. J., Verbi, W. & Crumpton, M. J. *EMBO J.* **5**, 1799–1808 (1986).
27. Gold, D. P. *et al. Nature* **321**, 431–434 (1986).
28. Hercend, T., Meuer, S., Reinherz, E. L., Schlossman, S. F. & Ritz, J. *J. Immun.* **129**, 1299–1306 (1982).
29. Reinherz, E. L. *et al. Cell* **30**, 735–743 (1982).
30. Transy, C., Moingeon, P. E., Marshall, B., Stebbins, C. & Reinherz, E. L. *Eur. J. Immun.* (in the press).
31. Hercend, T. *et al. J. clin. Invest.* **75**, 932–943 (1985).
32. Fiskum, G., Craig, S. W., Decker, G. L. & Lehninger, A. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3430–3433 (1980).
33. Rinnooy Kan, E. A., Wang, C. Y., Wang, L. C. & Evans, R. L. *J. Immun.* **131**, 536–539 (1983).

ACKNOWLEDGEMENTS. We thank Drs Michel Streuli, Haruo Saito and Marie-Luise Blue for advice and Carmaline O'Brien for technical assistance. This work was supported by grants from the N.I.H. P.A. is the recipient of a Regina S. Loeb award from the Arthritis Foundation. M.C. is the recipient of a grant from the Claudia Adams Barr Endowment for Cancer Research.

The LDL-receptor-related protein, LRP, is an apolipoprotein E-binding protein

Ulrike Beisiegel, Wilfried Weber, Gudrun Ihrke,
Joachim Herz*† & Keith K. Stanley*†

Medical Clinic, University Hospital Eppendorf, Martinistrasse 52,
D-2000 Hamburg 20, FRG
* European Molecular Biology Laboratory, Postfach 10.2209,
D-6900 Heidelberg, FRG

THE low-density-lipoprotein (LDL) receptor is a cell-surface protein that plays an important part in the metabolism of cholesterol by mediating the uptake of LDL from plasma into cells¹. Although LDL particles bind to the LDL receptor through their apolipoprotein B (apo B) and apolipoprotein E (apo E) moieties, other apo E-containing particles, like chylomicron remnants, are not dependent on the LDL receptor for uptake into cells. Chylomicrons formed in the intestinal mucosa during the absorption of the products of digestion, are processed by the peripheral circulation by lipoprotein lipase, which catalyses the breakdown of triglycerides in chylomicrons to free fatty acids and glycerol. The resulting chylomicron remnants, which are cholesterol-rich lipoproteins, are subsequently taken up in the liver^{2–6}. A second distinct protein that binds to apo E-containing lipoproteins, but not to LDL, has been proposed to be the receptor mediating the clearance of chylomicron remnants from the plasma. This protein has a relative molecular mass (M_r) of 56,000 (56K)⁷. More recent studies have failed, however, to establish whether this protein is a cell-surface receptor⁸. Here we describe crosslinking experiments in which apo E liposomes were found to bind specifically to the

cell surface of hepG2 cells and to human liver membranes. The size and immunological cross-reactivity of the protein to which the liposomes bound was indistinguishable from that of the recently cloned and sequenced LDL-receptor-related protein, LRP⁹. We therefore conclude that the LRP might function as an apo E receptor.

Several lines of evidence exist for the presence of a second distinct receptor that binds apo E-containing chylomicron remnants in the liver^{2–6}. In particular, the uptake of chylomicron remnants that contain apo E remains normal in animals with defective LDL receptors¹⁰. It is extremely difficult, however, to demonstrate a specific apo E receptor because the LDL receptor also binds to apo E with an even higher affinity than it does to the apo B of high relative molecular mass, apo B-100 (ref. 11). To differentiate between the receptors, we have previously used the approach of ligand blotting. Proteins from crude membrane fractions of human liver were electrophoretically separated, transferred to nitrocellulose filters and incubated with apo E-containing lipoproteins or apo E liposomes. With this method we detected only apo E-binding proteins of M_r 55–58K. Two proteins in this band were further analysed and found to be intracellular enzymes⁸. We report here the results of an alternative approach in which labelled apo E liposomes were bound and covalently crosslinked to apo E-binding proteins on intact cells. After subsequent solubilization we identified the major binding activity in a band corresponding to an M_r of ~600K that was indistinguishable from the band for LRP.

The first series of experiments were done with an ¹²⁵I-labelled heterobifunctional crosslinking agent that was coupled to apo E 3/4 by means of an amide linkage ($N[4-(p\text{-azido-}m\text{-}^{125}\text{I})\text{iodo-phenylazo}]benzoyl\text{-}3\text{-aminopropyl-}N'\text{-oxysuccinimide}$ ester; refs 12 and 13). After incubating labelled apo E liposomes with intact cells, unbound ligand was removed and covalent linkage induced by photoactivation. Subsequent cleavage of the intramolecular azo linkage with sodium dithionite leaves the ¹²⁵I-label on the binding protein rather than on the ligand. For these experiments we chose HepG2 (a human hepatocyte line) and P388 (a mouse macrophage line) cells because these cells

† Present addresses: Department of Molecular Genetics and Internal Medicine, UTHSCD, 5323 Harry Hines Boulevard, Dallas, Texas 75235, USA (J.H.) and the Heart Research Institute, 145 Missenden Road, Camperdown, New South Wales 2050, Sydney, Australia (K.K.S.).

contain much higher amounts of LRP than of the LDL receptor, especially if the cells are grown in fetal calf serum, when the LDL receptor is down-regulated. Figure 1 shows that the principal apo E-binding protein was very large, with an estimated M_r of $\sim 600K$. Autoradiography of the gels after transfer of the proteins to nitrocellulose filters enabled the mobility of the LRP to be determined on the same filter by western blotting. Although the bands were weak, the most heavily stained band coincided with the radioactively labelled apo E-binding protein (data not shown).

Because the LRP has a predicted M_r of 503K and contains many possible glycosylation sites⁹, we investigated the possible identity of the $\sim 600K$ apo E-binding protein with the LRP using crosslinking experiments on metabolically labelled cells. In this case a photoreactive heterobifunctional crosslinker was used (*N*-succinimidyl (4-azidophenyl)1,3'-dithiopropionate, SADP), which is cleavable with dithiothreitol (DTT). The experiments were performed as follows: HepG2 cells were preincubated for 4 h with [³⁵S]methionine to allow metabolic labelling of cellular proteins. After binding and cross-linking of apo E, the cells were solubilized with Nonidet P40. The lysates were then immune-precipitated in parallel with polyclonal antisera raised against apo E or the cytoplasmic tail of the LRP. Before

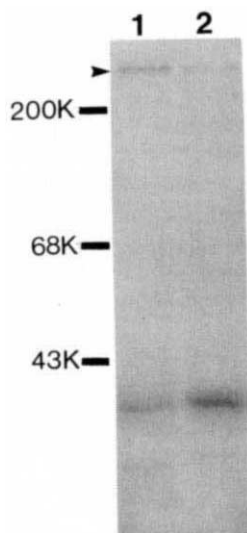


FIG. 1 Crosslinking of apo E liposomes to a surface protein of $M_r \sim 600K$ on HepG2 cells (lane 1) and mouse macrophages (P-388, lane 2). Reconstituted apo E 3/4 liposomes¹⁵ were coupled to ¹²⁵I-iodinated Denny Jaffe reagent and incubated with cells in lipid-free medium. After photoactivation and cleavage of the crosslinker, the cells were solubilized and comparable amounts of radioactivity were applied to an 8% SDS-polyacrylamide gel¹⁶. Proteins were transferred to nitrocellulose by electroblotting and autoradiographed for 17 days. An arrow marks the crosslinked band corresponding to an M_r of $\sim 600K$ which is radiolabelled; the band corresponding to an M_r of 34K represents the residual apo E-Denny Jaffe complex.

METHODS. Apo E 3/4 (220 μg) in sodium borate buffer (1 ml, pH 8) containing sodium cholate (23.4 mg) was added to a lipid film of egg-lecithin (2 mg) and cholesterol (40 μg). The mixture was vortexed for 1 min, gently shaken for 2 h at 24 °C, and finally dialysed against borate buffer, pH 8, for 48 h at 4 °C. The liposomes were isolated on a Sepharose CL 6B column as a homogenous peak and checked on an SDS-polyacrylamide gel for intact apo E. These proteoliposomes (15 μg) were coupled to 0.1 nmol (200 μCi) of Denny Jaffe reagent (NEN, NEX 227) according to the manufacturer's instructions. Cell monolayers ($\sim 10^7$ cells) were cultivated in DMEM containing 10% fetal calf serum and maintained in serum-free medium for 30 min before the experiment began to avoid direct lipoprotein contact. For the binding reaction the cells were incubated for 2 h at 4 °C with the apo E-Denny Jaffe complex. After the washing steps, photoactivation was with 366 nm UV light for 8 min. To cleave the linkage, the cells were treated with 200 mM dithionite. The cells were then solubilized in 200 μl Tris-HCl (50 mM, pH 8) containing 2% Triton X-100. The insoluble material was spun down for 10 min at 100,000 r.p.m. in a Beckmann TL-100 ultracentrifuge before PAGE.

electrophoresis, the immune-precipitates were reduced with DTT to separate the apo E from its cell-surface binding protein and restore it to its true M_r . The results in Fig. 2 show that the apo E binds to a band corresponding to an M_r of $\sim 600K$, which is indistinguishable from the band for LRP. The specificity of the LRP immune-precipitation was shown by competition with the peptide against which the antibody was raised (Fig. 2b, lane 3). The anti-apo E antibody did not immune-precipitate as much protein as the anti-LRP serum because at 4 °C apo E liposomes could only bind to the fraction of LRP on the cell surface, whereas the anti-LRP incubated with cell lysate could bind to all the LRP. Only a proportion of the total LRP was present on the cell surface at any one time (D. Murfitt and K.K.S., manuscript in preparation).

In a third series of crosslinking experiments, we looked for specific competition between the binding of apo E-containing liposomes and triglyceride-rich remnants (derived from very-low-

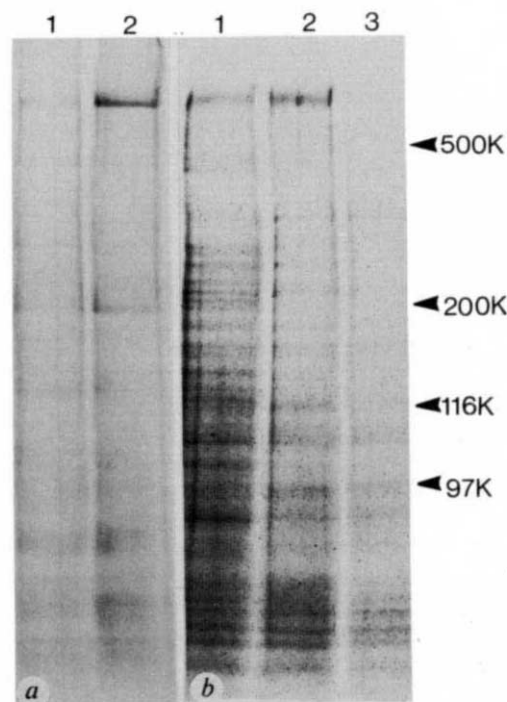


FIG. 2 Apo E-binding protein and LRP are indistinguishable. Metabolically labelled HepG2 cells were crosslinked to apo E liposomes using SADP (Pierce) and then immunoprecipitated using either anti-apo E or anti-LRP antisera. Before loading on SDS-polyacrylamide gels, the crosslinked proteoliposomes were dissociated in DTT. Two separate experiments are shown (a and b). Lane 1, polyclonal anti-apo E; lane 2, polyclonal anti-LRP serum; lane 3 (in b only), polyclonal anti-LRP serum in the presence of the peptide antigen against which the antibody was raised⁹ showing that the band corresponding to an M_r of 600K is specific.

METHODS. HepG2 cells in Petri dishes (10 cm) were metabolically labelled with [³⁵S]methionine (200 μCi) for 4 h. Apo protein E 2/2 (50 μg) reconstituted proteoliposomes (as described in Fig. 1) were coupled to the crosslinker SADP and incubated for 2 h at 4 °C with the ³⁵S-labelled HepG2 cells in Earle's salt solution (EBSS: Gibco). After washing, the bound apo E was crosslinked by photoactivation at 254 nm for 15 min and the cells then solubilized in Tris HCl (20 mM, pH 7.5), NaCl (150 mM), MgCl₂ (2 mM) containing 1% Nonidet P40. Solubilized cells from one Petri dish were split and incubated with 50 μl anti-LRP serum or 50 μl anti-apo E serum overnight. For peptide competition, 10 μg peptide was added to the antibody dilution before addition to the lysate. Immune complexes were immunoprecipitated by addition of 200 μl of a 1:1 suspension of protein A-Sepharose in the same buffer and incubation for 2 h at 4 °C. The samples were then centrifuged, and after extensive washing of the immunoprecipitate, the crosslinker was cleaved by addition of DTT (20 mM) and 5% SDS. The protein A-Sepharose was separated by centrifugation and a third of the supernatant loaded on a 6% SDS-polyacrylamide gel.

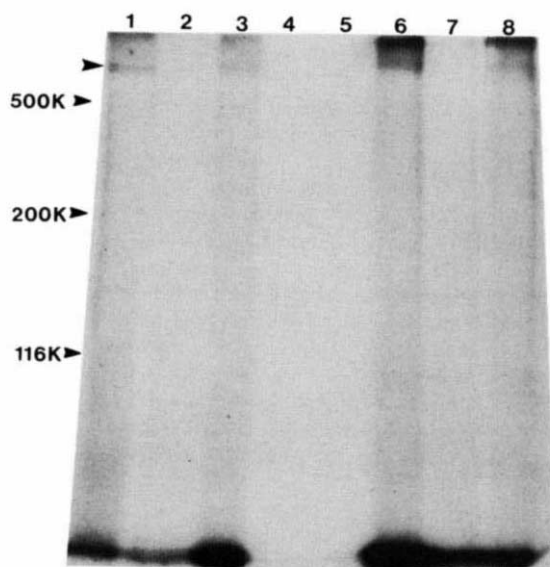


FIG. 3 Triglyceride-rich remnants compete for apo E binding to the LRP on HepG2 cells and human liver membranes¹⁷. About 10^7 cells (lanes 1–4) or 400 μ g membrane protein (lanes 5–8) were incubated for 1 h at 4 °C with 17 μ g of 125 I-iodinated apo E 3/4 liposomes (specific activity of 29.5×10^4 c.p.m. μ g⁻¹) either alone (lanes 1 and 6), in the presence of excess triglyceride-rich remnants (425 μ g protein; lanes 2 and 7) or with the addition of 2 mM EDTA (lanes 3 and 8). Lanes 4 and 5 represent control cells that received no apo E liposomes. Bound apo E was crosslinked to the cells after washing using disuccinimidyl suberate (DSS; Pierce) solubilized as described in Fig. 2 and loaded on a 6% SDS-polyacrylamide gel. Proteins were transferred to nitrocellulose and autoradiographed for 4 days.

METHODS. Triglyceride-rich remnants (apo E isotype 3/4) were isolated by density gradient centrifugation from a patient with remnant accumulation. Purified apo E 3/4 was iodinated using iodogen¹⁸ and reconstituted into liposomes as described in Fig. 1. Crosslinking was performed on washed cells or membranes using DSS in dimethyl sulphoxide at a final concentration of 1 mM for 30 min at room temperature.

density lipoproteins and chylomicrons). We also tested whether binding was sensitive to inhibition by EDTA. Because the shift in M_r did not matter in these experiments, we used a non-cleavable crosslinker (disuccinimidyl suberate). The results are shown in Fig. 3. The decoration of the band corresponding to an M_r of about 600K by 125 I-labelled apo E 3/4 liposomes was inhibited by incubation in the presence of excess unlabelled triglyceride-rich remnants when HepG2 cells or human liver membranes were used as the protein source (lanes 1 and 6 compared with lanes 2 and 7). With HepG2 cells, but not with the liver membranes, EDTA caused a decrease in, but not total inhibition of binding (Fig. 3, lanes 3 and 8). This indicates that the binding of apo E to LRP is not dependent on divalent cations, despite the strong binding of calcium ions to detergent-treated LRP on blots⁹. Because comparable amounts of LRP were present in all the lanes, as determined by immunoblotting of the same filter with anti-LRP antibody (data not shown), we conclude that the binding of apo E liposome is competitively inhibited by lipoprotein particles that contain apo E. Definitive studies on the binding properties of LRP must await biochemical purification of the protein.

We have shown using three different types of crosslinking reagents that the protein to which apo E liposomes bind in HepG2 cells, P388 cells and human liver membranes is indistinguishable from LRP, based on M_r and immunological criteria. Both apo E 2/2 and apo E 3/4, which contain all three apo E isoforms, are recognized by the LRP in a manner competed by natural apo E-containing particles. The binding of apo E 2/2 to the LRP could explain why patients homozygous for apo E 2/2 do not necessarily have elevated lipid levels, despite the inability

of remnants containing apo E 2/2 to bind to the LDL receptor¹⁴. Binding is also relatively independent of added EDTA, as would be expected for the chylomicron-remnant receptor⁴. It remains to be explained, however, why LRP is so widely distributed among tissues⁹ if its function is solely to bind chylomicron remnants. □

Received 30 May; accepted 24 July 1989.

1. Brown, M. S. & Goldstein, J. L. *Science* **232**, 34–47 (1986).
2. Sherrill, B., Innerarity, T. L. & Mahley, R. W. *J. biol. Chem.* **255**, 1804–1807 (1980).
3. Hui, D. J., Innerarity, T. L. & Mahley, R. W. *J. biol. Chem.* **256**, 5646–5655 (1981).
4. Cooper, A. D., Erickson, S. K., Nutik, R. & Shrewsbury, M. A. *J. Lipid Res.* **23**, 42–52 (1982).
5. Brown, M. S. & Goldstein, J. L. *J. clin. Invest.* **72**, 743–747 (1983).
6. Hoeg, J. M., Demosky, S. J., Gregg, R. E., Schafer, E. J. & Brewer, B. *Science* **227**, 759–761 (1985).
7. Hui, D. Y., Brecht, W. J., Hall, E. A., Innerarity, T. L. & Mahley, R. W. *J. biol. Chem.* **261**, 4256–4276 (1986).
8. Beisiegel, U. *et al. Arteriosclerosis* **8**, 288–297 (1988).
9. Herz, J. *et al. EMBO J.* **7**, 4119–4127 (1988).
10. Kita, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 3623–3627 (1982).
11. Pitas, R. E., Innerarity, T. L., Arnold, K. S. & Mahley, R. W. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2311–2315 (1979).
12. Denny, J. B. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5286–5290 (1984).
13. Jaffe, C. L., Lis, H. & Sharon, N. *Biochemistry* **19**, 4423–4429 (1980).
14. Utermann, G., Paulin, N. & Steinmetz, A. *Clin. Genet.* **15**, 63–72 (1979).
15. Helenius, A., Fries, E. & Kartenbeck, J. *J. Cell Biol.* **75**, 866–880 (1977).
16. Neville, D. M. *J. biol. Chem.* **246**, 6328–6334 (1971).
17. Kovanen, P. T., Basu, S. K., Goldstein, J. L. & Brown, M. S. *Endocrinology* **104**, 610–616 (1979).
18. Fraker, P. J. & Speck, J. C. *Biochem. biophys. Res. Commun.* **80**, 849–857 (1978).

The codon CUG is read as serine in an asporogenic yeast *Candida cylindracea*

Yoshiyuki Kawaguchi, Hiroshi Honda, Junko Taniguchi-Morimura & Shinjiro Iwasaki

Tokyo Research Laboratory, Meito Sangyo Co. Ltd., 2973-2 Ishikawacho, Hachioji, Tokyo 192, Japan

DEVIATIONS from the universal genetic code have been reported for several microorganisms. Termination codons are used for coding some amino acids in *Paramecium*^{1,2}, *Mycoplasma*³ or *Tetrahymena*^{4,5}, and in *Escherichia coli*, the UGA termination codon is used to code for selenocysteine⁶. In mitochondria, the changes of sense codons to termination codons or to codons encoding other amino acids have also been reported^{7,8}. Here we report another example of divergence from the universal code, this time in a non-spore-forming yeast *Candida cylindracea*, in which the universal codon for leucine, CUG, is used to code for serine. This conclusion is based on the observations that: (1) the amino-acid composition and the partial amino-acid sequences of an extracellular lipase from this yeast agreed with those deduced from the complementary DNA if CUG was assumed to specify serine; and (2) serine, but not leucine, was incorporated into a polypeptide in a cell-free translation system from this yeast in the presence of a synthetic CUG oligomer.

*Candida cylindracea*⁹ strain MS-5 (deposited as *Candida rugosa* ATCC14830) produces two kinds of extracellular lipase—lipase I of M_r 60,000 is the main component and lipase II (M_r 60.5K) is the minor component¹⁰. From this strain we isolated a mutant that produced a large amount of lipase I and a negligible amount of lipase II. We constructed a cDNA library from this mutant, from which a clone carrying the longest lipase I cDNA was obtained. This clone was named λ CIL15 (see Fig. 1 legend). We sequenced this cDNA and deduced its corresponding 534-amino-acid sequence which contained a sequence identical to that of the N-terminal portion (residues 1–20) of lipase I (Fig. 1). Part of the putative signal sequence was found upstream of the N terminus. Other cDNA clones with shorter inserts were also sequenced. Their sequences coincided almost exactly with those of the corresponding parts of the λ CIL15

FIG. 1 Nucleotide sequence of the cDNA from λ CL115 and its deduced amino-acid sequence according to the universal code. N terminus of lipase I is indicated by an arrow. The determined amino-acid sequences of the N terminus of lipase I and the portions corresponding to the sequences of the peptides of the lysyl endopeptidase digests (see Fig. 2 legend) are underlined. Examples of nucleotide and amino-acid substitution sites are boxed. CTG codons are indicated by asterisks. METHODS. Poly(A)⁺RNA from the *C. cylindracea* mutant that was cultured in medium containing 2% soybean oil was extracted by the guanidium/caesium chloride method¹⁶ followed by chromatography on oligo(dT)-cellulose. Using this poly(A)⁺RNA as a template, a cDNA library was prepared using λ gt11¹⁷ as a vector. From this library, clones carrying the partial cDNA sequences of lipase I were selected by immunodetection with anti-lipase I antibody according to the method described by Young and Davis¹⁸. Clones

carrying the longer cDNAs were screened by plaque hybridization with the above cDNA as a probe. The clone carrying the longest cDNA was named λ CL115. Complementary DNAs were sequenced by the dideoxynucleotide chain-termination method¹⁹.

1 Peptide	Phe Thr Ser Tyr Gly Pro Ser
cDNA	Phe Thr <u>Leu</u> Tyr Gly Pro <u>Leu</u>
	TTC ACG CTG TAC GGC CCG CTG
	169
2 Peptide	Val Phe Glu Ala Val Ser Pro Ser Ser Glu
cDNA	Val Phe Glu Ala Val <u>Leu</u> Pro <u>Leu</u> Ser Glu
	GTG TTT GAG GCG GTG CTG CCG CTG AGC GAG
	268
3 Peptide	Arg Ile Ser Ala Val Leu Gly Asp Leu Gly
cDNA	Arg Ile <u>Leu</u> Ala Val Leu Gly Asp Leu Gly
	AGA ATC CTG GCG GTG CTC GGC GAC CTT GGC
	1225

FIG. 2 Amino-acid sequences of the peptides from lipase I. Purified lipase I was digested with lysyl endopeptidase (*Acromobacter* protease I), and the peptides were separated by reverse-phase HPLC using a μ Bondapak C18 column (Waters). The N-terminal amino-acid sequences of the five peptides were determined with a gas-phase peptide sequencer (Applied Biosystems) (Fig. 1, underlined). The three peptides whose sequences disagreed with the cDNA sequence are illustrated. The upper line is the sequence derived from the peptide and the lower is that from the cDNA. Leucine residues corresponding to CTG are underlined.

cDNA apart from containing a few nucleotide substitutions, some of which caused substitution of amino acids (Fig. 1). These results indicate that in *C. cylindracea*, homologous genes encode several homologous lipase I proteins, which differ by a few amino-acid substitutions.

The amino-acid composition of purified lipase I was consistent with that deduced from the cDNA sequence, with the exception of the contents of leucine and serine, which were found to differ significantly.

TCGGTGGCTGCTGCCCCACCGCCACGCTCGCCACGGGACACCATCACGGTCTCAACGCCATCATCAACGAGGCGTCTCTGGCATTCCCTTTGCCGAGCGCGGTGGGCAACCTC
SerValAlaAlaAlaProThrAlaThrLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeu
CGCTTCAAGGACCCGCTGCCCTACTCGGCTCGCTCGATGGCAGAAAGTTCACGGTGTACGGCCCGCTGTGCATGCAGCAGAACCCGAGGGCACCTACGAGGAGAACCTCCCAAGGCA
ArgPheLysAspProValProTyrSerGlySerLeuAspGlyGlnLysPheThrLeuAlaValLeuProLeuSerGluAspCysMetGlnGlnAsnProGluGlyThrTyrGluGluAlaAsnLeuProLysAla
GCGCTCGACTTGGTGATGCATCCAAAGTGTTTGGGCGGTGCTGCCGTGAGCAGGAGCTGTCTACCATCAACGTGGTGGCGCCGCCGACCAAGCGGGTGCCAACTCCCGGTG
AlaLeuAspLeuValMetGlnSerLysValPheGluAlaValLeuProLeuSerGluAspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAlaAsnLeuProVal
ATGCTCTGGATCTTTGGCGCGGGTTTGAAGTGGTGGCAGCACCTTCCCTCCCGCCAGATGATCACAAGAGCATTGCCATGGGCAAGCCCATCATCCAGTGAAGCTCAACTAC
MetLeuTrpIlePheGlyGlyGlyPheGluValGlyGlyThrSerThrPheProProAlaGlnMetIleThrLysSerIleAlaMetGlyLysProIleIleHisValSerValAsnThr
CGCGTGTCTCGTGGGGGTTCTGGCTGGCGACGAGATCAAGGCCGAGGCGAGTGCACACCGCGTGAAGGACACGCGCTTGGCGATCGATGGTGGCGACAACATTGGCGCGTTT
ArgValSerSerTrpGlyPheLeuAlaGlyAspGluIleLysAlaGluGlySerAlaAsnAlaGlyLeuLysAspGlnArgLeuMetGlnTrpValAlaAspAsnIleAlaAlaPhe
GGCGGCGACCCGACCAAGGTGACCATCTTGGCGAGCTGGCGGCGACGATCGGTGATGCCACATCTCTGGAACGACGGCGACAACAGTACAAGGCAAGCGCTCTTCGCGCG
GlyGlyAspProThrLysValThrIlePheGlyGluLeuAlaGlySerMetSerValMetCysHisIleLeuTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAla
GGCATCATGCAGCTGGGGCCATGGTGGCTGGACCGCTGGACGCGATCTACGGCAACGAGATCTTGGCTCTTGGCGTGAAGCGGGCTGGCGAGCGCCAGCGACAAGCTTGG
GlyIleMetGlnLeuGlyAlaMetValProLeuAspAlaValAspGlyIleTyrGlyAsnGluIlePheAspLeuLeuAlaSerAsnAlaGlyCysGlySerAlaSerAspLysLeuAla
TGCTTGGCGGTGTGCTGAGCGACAGTGGAGGACGCCACCAACACCCCTGGGTCTTGGCGTACTCTCTGCTGGCTGTGTACCTCCCGCGCCGACGCGTGAACATCACC
CysLeuArgGlyValLeuSerAspThrLeuGluAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuLeuLeuProArgProAspGlyValAsnIleThr
GACGACATGTACGCTTGGTGGCGGAGGGCAAGTATGCCAATCCCTGTGATCATCGCGACCAAGACGAGGGCACCTTCTTGGACCCCTGCTGTGAACGTGACCCGATGCC
AspAspMetTyrAlaLeuValArgGluGlyLysTyrAlaAsnIleProValIleIleGlyAspGlnAsnAspGluGlyThrPhePheGlyThrLeuLeuLeuAsnValThrThrAla
CAGGCCCGGAGTACTTCAAGCAGCTGTTTGTCCAGCCGAGGCGAGGATGCACGCTGTATGACGGCGTACCCGAGCATCCCCGAGATCCCCGAGGCGCTCCGCTTCGACCGGATCTCT
GlnAlaArgGluTyrPheLysGlnLeuPheValHisAlaSerAspAlaGluIleAspThrLeuMetThrAlaTyrProGlyAspIleThrGlnGlyLeuProGlyAspThrGlyIleLeu
AACGCCCTCACCCGCGAGTCAAGAGAATCTGGCGGTGCTGGCGACCTTGGCTTACGGTGTCTGCTACTTCTCAACCATACACCGCGGACCAAGTACTCTCTCTCTG
AsnAlaLeuThrProGlnPheLysArgIleLeuAlaValLeuGlyAspLeuGlyPheThrLeuAlaArgArgTyrPheLeuAsnHisTyrThrGlyGlyThrLysTyrSerPheLeuLeu
AAGCAGCTCTGGGTTTGGCGGTGCTCGGAACCTTCCATCCACGACATGTCTTCCAGGACTACTTGTGGGACGGCTCGCTCATACAAACACGCGTTCATTGCGTTTGGCAG
LysGlnLeuLeuGlyLeuProValLeuGlyThrPheHisSerAsnAspIleValPheGlnAspTyrLeuLeuGlySerGlySerLeuIleTyrAsnAsnAlaPheIleAlaPheAlaThr
GACTTGGACCCCAACACCGCGGGTGTGGTGAAGTGGCCGAGTACACGACGCTCAGCTGGGCAACAACCTGATGATGATCAACGCTTGGGCTGTACACCGGCAAGGACAAC
AspLeuAspProAsnThrAlaGlyLeuLeuValLysTrpProGluTyrThrSerSerLeuGlnLeuGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsn
TTCCGACCGCGCGCTACGACGCGTGTCTTCCAAACCGCGCTTCTTGTGAAGGACCGTGTTCATATCTCAACAACGCTGTATCCACGCAAAAAA
PheArgThrAlaGlyTyrAspAlaLeuPheSerAsnProProLeuPhePheVal***

TABLE 1 Codon usage in lipase I mRNA

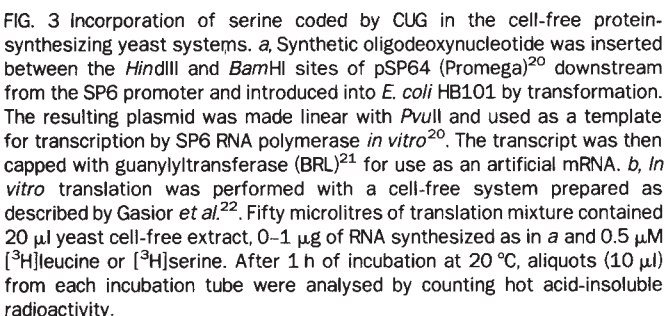
UUU	Phe	12	UCU	Ser	0	UAU	Tyr	1	UGU	Cys	1
UUC	Phe	19	UCC	Ser	5	UAC	Tyr	19	UGC	Cys	4
UUA	Leu	0	UCA	Ser	1	UAA	—	1	UGA	—	0
UUG	Leu	23	UCG	Ser	7	UAG	—	0	UGG	Trp	5
CUU	Leu	3	CCU	Pro	3	CAU	His	0	CGU	Arg	9
CUC	Leu	22	CCC	Pro	12	CAC	His	5	CGC	Arg	9
CUA	*	0	CCA	Pro	0	CAA	Gln	0	CGA	Arg	0
CUG	Ser	19	CCG	Pro	16	CAG	Gln	16	CGG	Arg	3
AUU	Ile	7	ACU	Thr	0	AAU	Asn	0	AGU	Ser	1
AUC	Ile	22	ACC	Thr	25	AAC	Asn	32	AGC	Ser	12
AUA	Ile	0	ACA	Thr	0	AAA	Lys	0	AGA	Arg	1
AUG	Met	14	ACG	Thr	11	AAG	Lys	20	AGG	Arg	0
GUU	Val	0	GCU	Ala	2	GAU	Asp	2	GGU	Gly	6
GUC	Val	4	GCC	Ala	24	GAC	Asp	32	GGC	Gly	44
GUA	Val	0	GCA	Ala	1	GAA	Glu	0	GGA	Gly	1
GUG	Val	24	GCG	Ala	20	GAG	Glu	18	GGG	Gly	5

Numbers indicate the frequency with which the codons are used in the coding region of lipase I mRNA with the exception of the putative signal sequence. The numbers of Leu and Ser residues, calculated by assuming that CTG encodes Ser, are 48 and 45, respectively. Those determined from amino-acid analysis were 47 and 43, respectively. Asterisk denotes that the amino acid encoded by CUA is unknown. —, Stop codons.

To discover the reason for this discrepancy we sequenced the internal region of lipase I. Figure 2 shows the partial amino-acid sequences of the peptides isolated from lysyl endopeptidase digests of lipase I and the corresponding sequences deduced from the cDNA. At all of the sites in the peptides encoded by CTG, leucine was substituted by serine. When the amino-acid sequence corresponding to the cDNA was re-deduced by regarding CTG as a codon specifying serine, it showed good agreement with the contents of leucine and serine found in the protein

To obtain direct evidence for the coding of serine by CUG, we performed an *in vitro* translation experiment with a synthetic CUG oligomer prepared as described in Fig. 3a. This artificial messenger RNA was introduced into cell-free translation systems from cells of either the mutant *C. cylindracea* or *Saccharomyces cerevisiae* JCM7255, and the incorporation of [³H]serine or [³H]leucine was then monitored (Fig. 3b). An AUC pentamer was incorporated into the mRNA with the expectation that the resulting isoleucine pentamer structure would facilitate precipitation of the translation product by trichloroacetic acid. In the *S. cerevisiae* system [³H]leucine incor-

Another remarkable feature of the *C. cylindracea* genome is the existence of a large number of pseudogenes. Several such sequences previously analysed have more than 90% homology with the lipase I cDNA (data to be reported elsewhere). Codon alteration in mitochondria or *Mycoplasma* is thought to be potentially useful for economic use of limited amounts of genetic material¹⁵. Our results, however, indicate that such abnormal codon usage can take place even in organisms possessing a genome with highly repetitive sequences. It is unlikely that such a phenomenon occurs in only a single microbial strain, and it would be interesting to clarify the distribution of similar codon alterations in other microorganisms. □



1. Caron, F. & Meyer, E. *Nature* **314**, 185-188 (1985).
2. Preer, J. R. Jr, Preer, L. B., Rudman, B. M. & Barnett, A. J. *Nature* **314**, 188-190 (1985).
3. Yamao, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 2306-2309 (1985).
4. Horowitz, S. & Gorovsky, M. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2452-2455 (1985).
5. Kuchino, Y., Hanyu, N., Tashiro, F. & Nishimura, S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4758-4762 (1985).
6. Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M.-A. & Böck, A. *Nature* **331**, 723-725 (1988).
7. Bonitz, S. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3167-3170 (1980).
8. Anderson, S. *et al. Nature* **290**, 457-465 (1981).
9. Yamada, K. & Machida, H. *Nippon Nōeigakagaku kaishi* **36**, 858-860 (1962).
10. Tomizuka, N., Ota, Y. & Yamada, K. *Agric. Biol. Chem.* **30**, 576-584 (1966).
11. Normanly, J., Ogden, R. C., Horvath, S. J. & Abelson, J. *Nature* **321**, 213-219 (1986).
12. Inokuchi, H., Höben, P., Yamao, F., Ozeki, H. & Söll, D. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5076-5080 (1984).
13. Gorini, L. in *Ribosomes*, 791-803 (Cold Spring Harbor Laboratory, New York, 1974).
14. Yarus, M. & Thompson, R. in *Gene Function in Prokaryotes* (eds Beckwith, J., Davies, J. & Gallant, J. A.) 23-63 (Cold Spring Harbor Laboratory, New York, 1983).
15. Osawa, S. & Jukes, T. H. *Trends Genet.* **4**, 191-198 (1988).
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning: A Laboratory Manual* 196 (Cold Spring Harbor Laboratory, New York, 1982).
17. Young, R. A. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1194-1198 (1983).
18. Young, R. A. & Davis, R. W. *Science* **222**, 778-782 (1983).
19. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
20. Melton, D. A. *et al. Nucleic Acids Res.* **12**, 7035-7056 (1984).
21. Krieg, P. A. & Melton, D. A. *Nucleic Acids Res.* **12**, 7057-7070 (1984).
22. Gasior, E., Herrera, F., Sadnik, I., McLaughlin, C. S. & Moldave, K. *J. biol. Chem.* **254**, 3965-3969 (1979).

NATURE · VOL 341 · 14 SEPTEMBER 1989

Scintillation proximity assay

N. Bosworth and P. Towers

Fluomicrospheres coated with antibodies or receptors eliminate the need to separate bound from free ligand in both radioimmunoassays and ligand-binding assays.

MANY advances have been achieved in radioimmunoassays (RIA) and ligand-binding assays (LBA) in recent years, but the physical separation of bound from free ligand still remains a time-consuming operation that cannot be automated and that often can lead to increased non-specific binding. Both RIA and LBA expose researchers to the hazards of manipulating radioactive solutions of sometimes unknown biological activity. Tritium-based assays require the use of liquid scintillant, which is flammable, toxic and expensive to purchase and dispose.

Amersham International has used scintillation proximity assay (SPA) systems¹⁻⁷ to overcome the problems outlined above. SPA involves the use of fluomicrospheres coated with acceptor molecules, such as antibodies or receptors, to which a ligand will bind selectively in a reversible manner.

The technique is carried out in normal assay buffers, and requires the use of a ligand labelled with an isotope that emits low-energy radiation which is dissipated easily into an aqueous medium. It is fortunate that two of the most widely used isotopes in life science research — ^3H and ^{125}I — possess ideal emission characteristics for their use in SPA.

Because ^3H β -particles and ^{125}I Auger electrons have average energies of 6 and 35 keV, respectively, their energies are absorbed by aqueous solutions within very small distances (4 μm for ^3H β -particles and 35 μm for ^{125}I Auger electrons). The mean path length of ^{125}I gamma emission is 15 cm, so very little of the energy is absorbed by the fluomicrospheres or the buffer.

Bound and free

At any point during an assay, bound labelled ligands will be in close proximity to the fluomicrospheres, allowing the emitted energy to activate the fluor and produce light. Because the assay typically is carried out in 0.4 ml of buffer, the majority of unbound labelled ligands are too far from the fluomicrospheres, which are present in microgram quantities, to enable the transfer of energy. Bound ligands produce light, but free ligands do not, obviating the need for any of the traditional tedious separation procedures (Fig. 1).

The limited amount of published work concerning SPA has been carried out using fluomicrospheres based on organic phosphors, such as diphenyloxazole-latex⁶

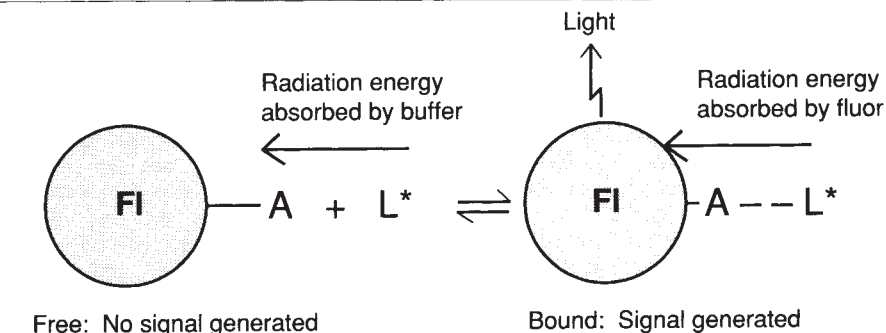


FIG. 1 The principle of scintillation proximity assay: only labelled ligand (L^*) that is bound to the acceptor molecule (A) generates a signal on the fluomicrosphere (FI).

and polyvinyltoluene plastic scintillator beads². Assays performed using bead fluor complexes prepared by adsorbing organic compounds into a matrix are potentially susceptible to error, because the fluor may leach into the buffer.

To increase the photon yield and remove the possibility of loss of fluor and subsequent assay drift, Amersham has developed a new type of derivatized inorganic fluomicrosphere based on yttrium silicate. The yttrium silicate is doped selectively with rare earth elements to facilitate the production of light with the optimum emission characteristics for the photomultipliers and electronic circuitry used in commercially available scintillation counters. When used in SPA, these inorganic fluomicrospheres produce up to twice the photon yield compared to any of the currently available organic fluor beads⁸.

Amersham has optimized the size of the fluomicrospheres for both RIA and LBA, by balancing a maximal surface area-to-weight ratio against the requirement that the radiation energy be absorbed by, rather than pass through, the fluomicrospheres. Spheres in the size range of 1–5 μm provide a large enough surface area to immobilize a sufficient number of acceptor molecules, and yield a high amount of photons⁸.

A range of synthetic methods has been developed to couple acceptor molecules and fluomicrospheres, because physical adsorption onto non-activated fluomicrospheres produces complexes which are not tightly bound and which have poor long-term stability. To overcome the resulting problems of variable assay reproducibility, activated fluomicrospheres (such as those bearing carboxyl, amino, epoxy and aldehyde groups) are used in the coating process. Acceptor-fluomicrosphere complexes coupled covalently using mild, non-denaturing conditions yield optimum

assay performance.

Two approaches have been used to convert conventional heterogeneous RIAs into an SPA format. Initially, primary antibodies were coupled directly to functionalized fluomicrospheres (for example, anti-bungarotoxin antibodies for use in a [^{125}I]Bungarotoxin assay). The method proved satisfactory, but incomplete coupling of valuable antibodies was a problem. Amersham now uses a second immunosorbent approach based on anti-rabbit, anti-sheep, anti-mouse, protein A and protein G fluomicrospheres. These have been used to convert over 20 ^{125}I and ^3H RIAs for a spectrum of compound types, for example, prostaglandins, leukotrienes, peptides and proteins. The single-tube assay is carried out by adding 0.1-ml quantities of the required reagents and allowing equilibrium to be reached before counting in a scintillation counter. In all of the assays that Amersham has performed, the results produced by the traditional RIA and SPA protocols have been identical.

Receptor studies

Amersham uses SPA technology to study solubilized and membrane-bound receptors. Soluble receptors are attached to fluomicrospheres by chemical bonding, lectin-glycoprotein interaction, or receptor/anti-receptor binding. Solubilized epidermal growth factor (EGF) receptors have been covalently bonded to aldehyde-derivatized fluomicrospheres. One drawback of this non-specific approach is the requirement for relatively pure preparations of receptor isolated from cell membranes.

Improved results are obtained when lectin fluomicrospheres, such as those bound to Concanavalin A or wheat germ agglutinin, are used⁹. The coupling of mouse anti-EGF receptor monoclonal antibody (MAb) that does not affect EGF

ligand binding onto the anti-mouse antibody fluomicrospheres enables the production of saturation and displacement curves using [125 I]EGF and Triton X-100 solubilized post-nuclear supernatant cell membrane preparations⁸. If any of the reagents in the complex (fluomicrosphere — anti-mouse antibody — anti-EGF receptor MAb — EGF receptor — [125 I]EGF) are omitted, no binding is detected, suggesting the bound complex has the above structure.

Because Scatchard analysis requires bound and free ligand to be in the same unit of measurement, [125 I] fluomicrospheres have been chemically synthesized as standards so that counts from gamma and scintillation counters can be inter-converted. The conversion factor is determined as the ratio of counts obtained using a gamma and a scintillation counter, and is found to be constant regardless of the specific activity, age or label:fluomicrosphere ratio of the standards. Using this conversion factor, the calculated dissociation constants are equivalent to those obtained by more traditional technology.

The coupling of membrane-bound receptors to fluomicrospheres allows the use of general methods that are not possible with solubilized receptors. Fluomicrospheres coated with entities that avidly bind one receptor type present in a membrane leave other receptors available for study (specific [125]transferrin binding is observed when A431 membranes bound to anti-EGF receptor MAb — anti-mouse antibody — fluomicrospheres are used⁸). Fluomicrospheres coated with antibodies to abundant and widely distributed membrane components, such as 5'-nucleotidase and Na⁺/K⁺ATPase are employed to study receptors in a variety of cell types, such as A431, KB, C6, rat lung and human placenta.

When these generic fluomicrospheres are used to investigate the β -adrenergic receptor in C6 cell membranes, the correct displacement profile is obtained using [125]iodocyanopindolol and (+)propranolol, (–)propranolol, (+)isoproterenol, (–)isoproterenol and (–)epinephrine. In a similar manner, LBAs have been developed for a range of receptor types, such as EGF, fibroblast growth factor, nerve growth factor, platelet-derived growth factor, endothelin and transferrin.

The conversion of RIAs and LBAs to SPA methodology confers many unique and significant advantages. Assays can be performed more safely, with fewer 'hands-on' manipulations, and the use of liquid scintillant is not necessary. Costs are also lower because fewer consumables such as tubes and filters are required. The precision of the assays is improved because the only manipulation required is pipetting of reagents, and higher throughputs can be achieved because the tech-

Innovative immunology

Buffers in tablet form, a bright-blue fluorescent fluorophore, and protein G immobilized to membranes and gels are a selection of this week's new ideas for the immunologist.

AN ANTI-mouse monoclonal isotyping kit for **characterizing mouse monoclonal antibodies** in tissue culture supernatants and ascites is available from Bioproducts for Science, Inc. (*Reader Service No. 101*). The test is based on red cell agglutination technology, and uses a rat MAb specific for a single class/subclass of mouse immunoglobulin that has been coupled to sheep red blood cells. The specific isotyping reagent is mixed with the supernatant under investigation and left for one hour. A positive result (agglutination) is indicated by lattice formation in the well of the plate, and a negative result by "button" formation (no agglutination). The test can detect antibody in the 1–20 μ l per ml range, says the company. The \$116.30 (US) kits are supplied with enough reagents for 30 tests and include freeze-dried anti-mouse IgG1, IgG2a, IgG2b, IgG3, IgA, and IgM reagent cells and control cells, reconstitution diluent and microtitre plates. Freeze-dried reagents can be stored for 6 months at 4 °C and for 2–3 months when reconstituted, says BPS.

ProAna Mabs is a new system for **monitoring monoclonal antibody production** from EDT Analytical Ltd (*Reader Service No. 102*). Using HPLAC technology developed by Biolytica, the system combines the selectivity of affinity chromatography with the speed and efficiency of HPLC. Central to the system is a patented Fc receptor that is specific for IgG. The £950 (UK) ProAna Mabs system comes as a kit containing a step-by-step manual, ProAna Mabs column, IgG standard, binding and eluting buffers, precolumn filter, 1-ml sample loop, and all the fittings necessary for rigging the column up to any HPLC machine. The sample is injected into the HPLC system, the column is

nology lends itself to automation.

The future of SPA technology is very exciting, because it solves many of the problems encountered when it is necessary to separate bound and free ligand. There are important concurrent developments in related research areas: antibody and receptor solid-surface coupling methods are being improved, cloned receptors (especially human) are becoming increasingly available, and inorganic phosphors are being developed with higher photon yields. The inevitable availability of more sophisticated scintillation counters will allow the simultaneous counting of many samples. □

washed and the antibodies are eluted: MAb concentration is determined by the UV detector — a procedure which EDT says takes 10 minutes. The ProAna Mab column can be used over 1,000 times without a reduction in yield, says EDT.

Using Sigma Chemical's **buffer tablets** for phosphate-buffered saline and Tris-buffered saline you can dispense with the need to weigh out compounds (*Reader Service No. 103*). Tableted buffers offer the advantages of convenience of use, accuracy of dosage, and improved stability over powdered forms, says Sigma.



Buffers in tablet form from Sigma.

Formulated to exacting specifications for use in immunoassays and enzyme assay procedures, the buffer tablets are available in quantities of 50 and 100 tablets for \$19.80 (US) and \$33 (US), respectively. Bulk quantities are also available.

Laboratory hardware

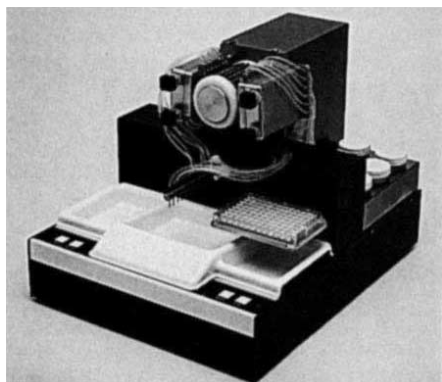
Beckman Instruments, Inc. offers its LS 6000SE **liquid scintillation counter** for use in RIA studies (*Reader Service No. 104*). Designed to reduce the number of operating decisions, the system features built-in help screens, preset counting regions,

Nigel Bosworth and Pat Towers are at Life Sciences Business Development Group, Amersham International plc, Cardiff Laboratories, Forest Farm, Whitchurch, Cardiff CF4 7YT, UK. For more information, fill in reader service number 100.

- Hart, H.E. & Greenwald, E.G. *Molec. Immun.* **16**, 265–267 (1979).
- Udenfriend, S., Gerber, L.D., Brink, L. & Spector, S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8672–8676 (1985).
- Nelson, N. *Analyt. Biochem.* **165**, 287–293 (1987).
- Udenfriend, S., Gerber, L.D. & Nelson, N. *Analyt. Biochem.* **161**, 494–500 (1987).
- Hart, H., US Patent No. 4,382,074 (1983).
- Hart, H., US Patent No. 4,388,296 (1983).
- Bertoglio-Matte, J.H., European Patent Application No. 84301769.0 (1984).
- Bosworth, N. & Towers, P. (in preparation).

and quench-correction curve setup parameters. New features include a coincidence monitor for determining non-radioactive events such as chemiluminescence and the Xtalscint option for calculating d.p.m.'s when solid scintillators are used. Beckman says less than 200 μ l of sample is needed for c.p.m. analysis. Up to ten user programs can be stored by the machine. The \$16,700 (US) LS 6000SE comes with an 80-column printer, sample racks, alphanumeric keypad, high-activity monitor and electrostatic controller.

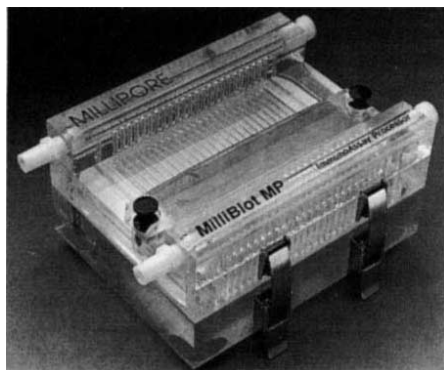
Where speed and high-throughput of **liquid dispensing** are important considerations, IBG Systems Ltd recommends its Inverness Dispenser (*Reader Service No. 105*). The £5,450 (UK) system can automatically dispense up to 12 different reagents simultaneously into a standard 96-well microtitre plate, in either the 8- or 12-row orientation. The machine takes less than eight seconds to load a microtitre



IBG Systems' 12-reagent dispenser for microtitre plates.

plate with 35 μ l per well, says IBG. Magnetic stirrers can keep up to six reagents or cell suspensions evenly mixed when the machine is dispensing. The operator simply primes the system each day or when the reagents are changed and selects the volumes to be dispensed per well, which can be in 5- μ l increments from 5 μ l to 500 μ l. Cross-contamination is avoided because the reagent paths are kept separate. The peristaltic pump has a reverse-control function that allows unused reagents to be returned to their bottles.

Millipore says its new MilliBlot-MP immunoassay processor is ideal for screening antibodies against a single antigen (*Reader Service No. 106*). The \$1,000 (US) multiple-channel assay device can process 1-24 channels. It uses a standard 7.5 cm $\times$ 8.5 cm single-blotted membrane sheet, making it unnecessary to cut and handle individual membrane strips, says Millipore. The device is fitted with a spring-loaded interlock which provides a compression fit and produces distinctly embossed channels on the membrane with a raised reaction surface. This, the com-



MilliBlot-MP: Millipore's new immunoassay processor for antibody screening.

pany says, gives precise chambering for consistent results from channel to channel and eliminates 'cross-talk'. As little as 100 μ l of serum is required per channel; antibodies can be detected from dried blood spots with volumes of 25-100 μ l, says Millipore. Conjugate, wash, and chromogen incubations can be carried out within the system, without transferring the membrane.

Designer labels

AMCA — 7-amino-4-methylcoumarin-3-acetate — is the new **bright-blue fluorophore** from Jackson Immunoresearch Laboratories for use in single and multiple labelling experiments (*Reader Service No. 107*). AMCA is now available conjugated to the company's AffiniPure and Affini-Sorbed antibodies, ChromPure immunoglobulin controls, streptavidin and avidin. The fluorophore absorbs light at 350 nm and fluoresces at 450 nm; this 100-nm Stokes shift produces minimal overlap between excitation and emission spectra, says the company. Jackson Immunoresearch Labs says AMCA's minimal overlap with other fluorophores makes it especially useful in the multiple colour analysis of single sections.

A novel range of **immunodetection reagents** that offer an alternative to streptavidin/avidin detection systems are available from British Bio-technology Ltd (*Reader Service No. 108*). BBL says its Ultra Hapten reagents are based on a high-affinity MAb to biotin and are capable of detecting any biotinylated target molecule. The MAb to biotin is available conjugated to a range of molecules, including alkaline phosphatase, fluorescein, gold, peroxidase, and Texas Red. The company says its reagents overcome the lack of sensitivity and nonspecific binding problems associated with streptavidin and avidin. Stated applications include immunocytochemistry, *in situ* hybridization, Western blotting and electron microscopy.

Immobilization chemistry

Everything you need to prepare **Fab and F(ab')₂ fragments** from IgG is now avail-

able in a kit from Pierce (*Reader Service No. 109*). Pierce's ImmunoPure Fab preparation kit uses immobilized papain to cleave the IgG; the ImmunoPure F(ab')₂ kit uses immobilized pepsin. After cleavage of the IgG, the fragments can be easily separated from the immobilized enzymes, eliminating the need for an ion-exchange separation step, says Pierce. The \$295 (US) kits also contain protein A columns and all the buffers required to mop up the Fc fragments and any undigested IgG, leaving purified fragments which are eluted in the void volume.

ADVERTISEMENTS

Watchful Habbits Keep Winning Rabbits

ASK ABOUT ANTIBODIES CUSTOMS
ESTABLISHED 1961

GOOD FEED, RELIABLE BREEDERS
FREQUENT VET INSPECTIONS
EAR BLEEDS, USDA INSPECTED
LET'S TALK OVER YOUR PROTOCOL TODAY.

800-824-8540

Antibodies Incorporated
SERVING SCIENCE THROUGH IMMUNOCHEMISTRY

Reader Service No.25

Custom DNA

Purified and Delivered in 48 hours

100 ug, \$10.00 a base for the first 15 bases, \$7.50 for each additional base. Milligram quantities available.

Research Genetics

2130 S. Memorial Parkway Huntsville, AL 35801

1-800-533-4363

In the United Kingdom 0-800-89-1393

In Canada 1-800-833-8348

FAX 205-536-9016 (NW2971)D

Reader Service No.15

Bioferm

- Gamma-Linolenic Acid
- Evening Primrose Oil
- Borago Officinalis Oil

4690 Maple Grove Road
BEAMSVILLE, Ontario
CANADA L0R 1B0

Phone: 416-562-4311

Tlx: 061-5105

Fax: 416-562-5658

Reader Service No.27

Genex says its **protein G affinity membranes** make IgG purification as easy as filtering a sample (*Reader Service No. 110*). The AbSorbent G affinity discs contain GammaBind G — recombinant protein G bound to a polymer membrane. Samples are loaded by attaching a syringe or pump to the disc; unwanted proteins are washed away and purified IgG is eluted at low pH. Costing between \$145–350 (US), the discs are available in three sizes, which Genex says can purify between 0.75 mg and 15 mg of IgG in less than 15 minutes. Genex says the stable ligand attachment and protease-resistance of GammaBind allow the discs to be



Protein G affinity membranes from Genex.

reused. The AbSorbent G discs can be used to purify IgG from dilute hybridoma supernatants and from mouse, human, rat, cow, horse, sheep, rabbit and goat IgG subclasses.

For the rapid purification of IgG, Zymed Laboratories, Inc. recommends its **prepacked protein G columns** (*Reader Service No. 111*). Zymed's \$105 (US) 5-cm column contains 1 ml of recombinant-protein G Sepharose 4B gel. The gel contains 2.5 mg of immobilized recombinant protein G, which the company says has a binding capacity of 20–22 mg of human IgG per run. High yields of bound IgG are produced after low-pH elution, says Zymed. The column can be used for the purification of IgG from serum, ascites or supernatant, or for the removal of unwanted IgGs from samples containing other Ig classes. Zymed says the columns can be used for at least 20 runs. For those who prefer to pack their own columns, the company offers recombinant protein G Sepharose gel in 2- and 5-ml quantities, for \$120 (US) and \$250 (US), respectively.

Affinichrom is the name of a new line of **immunoaffinity chromatography columns** from Kontes Biotechnology (*Reader Service No. 112*). The columns contain non-porous glass beads in sizes from 0.25 mm to 3 mm, coated with avidin, streptavidin, protein A or Concanavalin A. The beads do not exhibit the compression problems and diffusion limitations on flow rate associated with other packing materials, says Kontes. Antibody binding capacity is maximized because the protein-coated column packings bind to the Fc portion of the antibody and maintain proper orienta-

tion of the antigen receptor sites. The protein-coated glass beads come pre-packed in three sizes of stainless steel HPLC column and one size of glass column, costing between \$350 and \$1,000 (US). Kontes also offers bulk packings for protein A, avidin and Con A, and a custom column-packing service.

The single-step purification of mouse MAbs from ascites or cell culture supernatant can be achieved using **protein A-agarose** from Boehringer Mannheim (*Reader Service No. 113*). Recombinant protein A is used in the production of protein A-agarose, which Boehringer says is free of the mitogenic staphylococcal enterotoxins that contaminate protein isolated from *Staphylococcus aureus*. Immobilized protein A has a high IgG binding capacity and leakproof stability over a wide range of pH and buffer conditions, says the company. Depending on the elution conditions used, Boehringer says the agarose gel can be used over 30 times. Boehringer also offers recombinant protein G.

Radioimmunoassay kits

Three new ¹²⁵I RIA kits are available from Advanced Magnetics, Inc. for the measurement of interleukin-1 alpha, interleukin-1 beta and interleukin-2 (*Reader Service No. 114*). The IL-1 alpha and IL-1 beta kits are based on competitive immunoassays where the analyte in the sample competes with the radioactively labelled analyte for a limited number of antibody sites. The IL-2 kit is a solution-phase RIA based on the 'sandwich' principle, where the IL-2 becomes sandwiched between anti-IL-2 (rabbit) and ¹²⁵I monoclonal IL-2 tracer. The addition of magnetic particles coupled to goat anti-rabbit IgG antibody followed by magnetic separation or centrifugation produces a pellet which can then be counted in a gamma counter. The interleukin concentration is determined by interpolation from a standard curve. AMI says the IL-1 alpha, IL-1 beta and IL-2 kits have minimal cross-reactivities and can detect 3.0 pg per ml, 0.05 ng per ml and 0.15 units per ml, respectively. The kits have enough reagents to perform 100 assays.

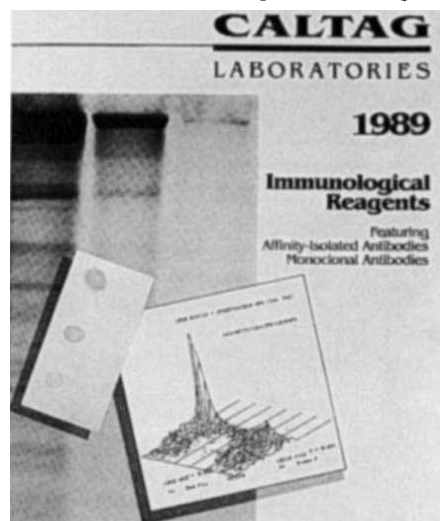


¹²⁵I RIA kits for measuring interleukin concentration, from AMI.

Collaborative Research, Inc. has a new line of **magnetic lymphokine RIA kits** for the detection and quantitation of IL-1 alpha, IL-1 beta, and IL-2 in culture media or biological fluids (*Reader Service No. 115*). The \$415 (US) ¹²⁵I Magnetic RIA Kits can detect picogram levels of both natural and recombinant lymphokines, says the company. The RIAs use a magnetic second antibody to effect the magnetic bench-top separation of the captured analyte. Results can be obtained in less than a day — half the time required for conventional bioassays, says Collaborative Research. Each kit contains a user's guide and all the reagents, standards and controls to perform 100 assays.

Biochemicals

Caltag Laboratories has a new **catalogue of immunological reagents** that lists over 300 monoclonal and affinity-isolated polyclonal antibodies — both unconjugated and conjugated to enzymes, fluorochromes, biotin and agarose (*Reader Service No. 116*). Caltag offers an expan-



Caltag Laboratories' catalogue of immunological reagents.

ded line of phycoerythrin and allophycocyanin conjugates for use in flow cytometry, new mouse monoclonals to BrdU/IdU, rat immunoglobulins, bacterial alkaline phosphatase and mouse Thy-1.2.

MonoCarb AB's product catalogue features a new range of **MAbs to carbohydrate epitopes** (*Reader Service No. 117*). The antibodies have important applications as a research tool in the fields of serology, oncology, and infectious disease research, but could also have diagnostic and therapeutic uses, says MonoCarb. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain additional information about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.